Department of English

Passive voices

BE-, GET- and prepositional passives
in recent American English

SARAH SCHWARZ

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Geijersalen, Engelska parken, Thunbergsvägen 3, Uppsala, Friday, 20 April 2018 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Anne Curzan (University of Michigan, Department of English).

Abstract
Schwarz, S. 2018. Passive voices. BE-, GET- and prepositional passives in recent American English. 67 pp. Uppsala: Department of English, Uppsala University. ISBN 978-91-506-2682-7.

The aim of the thesis is to shed light on the use and development of passive voice in American English. Empirical, corpus methods are employed in order to examine the syntactic, semantic, and stylistic preferences of three English passive constructions across time and genre in American English. The corpus data span the years 1870–2010 and come from genres of widely varying formality. The three passive constructions investigated in this thesis are:

1. The canonical BE-passive, as in *she was sent home*.
2. The informal, relatively infrequent GET-passive, as in *she got sent home*.
3. The typologically rare prepositional passive, as in *she was sent for*.

In Article 1, the frequency of BE- and GET-passives in very recent, speech-like material suggests both colloquialization and prescriptivism as influences on the language. The results indicate little difference between the two passives except in terms of frequency, highlighting the importance of comparing GET-passives to a control group of BE-passives. In Article 2, data from the TIME Magazine Corpus indicate that GET-passives may have been continuing to grammaticalize over the 20th century in terms of situation-type preferences. Article 3, which encompasses a longer diachronic span across more genres, lends further support to the continuing grammaticalization of GET-passives, and offers two additional indicators: decreased use with human subjects, and increasing acceptability with a range of past participles. Finally, the study of prepositional passives presented in Article 4 constitutes an empirical investigation of earlier theories against a control group of non-prepositional passives. The findings suggest diachronically stable differences along a range of features, including the thematic roles conferred on the passive subject-referent, supporting earlier claims about affectedness and perceptual salience of subject in prepositional passives.

The overall findings of the thesis highlight differences and similarities in three kinds of passive, and nuance our understanding of what passive voice is by using empirical methods to refine intuitive theories. The results regarding the use and development of the passives across time period and genre offer insight into the intertwined nature of mechanisms relating to language change, such as prescriptivism, colloquialization, and grammaticalization.

Keywords: passive voice, get-passive, prepositional passive, corpus linguistics, American English, historical syntax, adversativity, situation type, thematic role, affectedness, prescriptivism, colloquialization, grammaticalization

Sarah Schwarz, Department of English, Box 527, Uppsala University, SE-75120 Uppsala, Sweden.

ww

List of papers

This thesis is based on the following studies, which are referred to in the text by their Arabic numerals.

1. Schwarz, Sarah. 2015. Passive voice in American soap opera dialogue. *Studia Neophilologica* 87(2), 152–170.
2. Schwarz, Sarah. 2017. "Like getting nibbled to death by a duck": Grammaticalization of the GET-passive in the TIME Magazine Corpus. *English World-Wide* 38(3), 305–335.
3. Schwarz, Sarah. Under review a. Signs of grammaticalization: Tracking the GET-passive through COHA. Submitted to: fc. (2018). Claridge, Claudia & Birte Bös (eds.), *English Historical Linguistics 2016. Selected Papers from the Nineteenth International Conference on English Historical Linguistics (ICEHL 19)* (working title). Amsterdam & Philadelphia: Benjamins.
4. Schwarz, Sarah. Under review b. *This must be looked into*: A corpus study of the prepositional passive. Submitted to *Journal of English Linguistics*.

Reprints of Articles 1 and 2 were made with kind permission from the publishers; reprints of Articles 3 and 4 were made with kind permission from the editors of the volume and journal, respectively, to which the papers have been submitted, in accordance with the regulations of the respective publishers.

www

Acknowledgments

When I got accepted to the PhD program, I was more than a little concerned that doing research full time would be a very lonely endeavor. I'm so pleased at how wrong I turned out to be! As a PhD student, I rarely spent a lonely day, and have had so much support from supervisors, lecturers, fellow students, friends, and family that the years have flown by in a blur of pleasant activity. Research, as it turns out, is meant to be shared, and whether at higher seminars at the department of English, at outreach events for local high school students, or as a delegate at international conferences, it has been a joy to learn how to be an active member of a research community. This PhD was not a solitary effort by any means, and I find myself with a large number of people to thank for seeing me through it.

I would first like to thank my main supervisor, Professor Merja Kytö, who always allowed me to maintain a high degree of independence in my research while remaining constantly available for consultation and advice. A doctoral education involves a great deal more than the dissertation, and, with Merja's patient guidance, I have learned at least as much about what it means to be a professional researcher as I have about passive voice in American English. I would also like to thank my secondary supervisor, Docent Christer Geisler, who has provided valuable help and advice, especially with statistical testing and syntactic classification. I am so grateful for his genuine, infectious interest in English syntax. And also for having taught me how to use Excel Pivot Tables almost a decade ago. They were absolutely essential to the thesis!

Much gratitude also to all of the linguists who have read my drafts over the years at the department's higher research seminars. It has meant so much to me that you have taken the time to offer your feedback on my research, and your comments have enhanced the thesis greatly. Dr. Linnéa Anglemark, Dr. Gregory Garretson, Dr. Angela Hoffman, Dr. Christine Johansson, Dr. Ewa Jonsson, Dr. Henrik Kaatari, Dr. Tove Larsson, Dr. Pia Norell, Per Otterstedt, Dr. Göran Rönnerdal, Docent Erik Smitterberg, Erika Berglind Söderqvist, Professor Terry Walker, Dr. Ying Wang—thank you all so much for the time and effort you have devoted to helping me construct this thesis!

My warm thanks also go out to my fellow PhD students, past and present, at the Department of English. It has been such a pleasure working with you, and I thank you for all of your kindness and support. My office mates over

the years, Linnéa, Tove, Erika, and Sally, have all become dear friends. Though it's a cliché to say so, we have actually shared laughter and tears, and I feel very close to each of you.

To the kind of friend that you don't even think it's possible to make as an adult: thank you, Christine Mackay Tircomnicu, for your boundless support, understanding, and thirst for coffee! Thanks also to dear friends Karen Williams and Meena Strömqvist. You feel like family to me, and I'm so very happy to have you in my life.

My relatives and childhood friends live thousands of miles away, but they have all cheered me on from afar. My mother and father seemed to take my doing a PhD as a matter of course, which is astonishing considering how long it took me to get started. My grandmother, aunts and uncles, my brother, and my dear friends Natasha Wertheimer and Megan Tresham have believed I could do this even at times when I doubted it myself. Thank you all!

And to my Erik: thank you for all of your love and support. You are a brilliant linguist, but an even better partner – ja elska di, Island Boy.

And finally, thank you to my wonderful sons, who have had to put up with more linguistics conversation around the dinner table than any child should ever have to. I can't promise that I'll stop talking about linguistics in the future, but I *can* promise that you are allowed to tune me out with total impunity when I do. I love you both and am prouder of you than of anything else in my life.

Uppsala, January 2018

Sarah Schwarz

Contents

www

1. Introduction

In this compilation thesis, which comprises four articles (two published, two under review at the time of writing), material from some of the large, digitalized corpora freely available from Brigham Young University (BYU) (Davies 2007, 2010, 2012) was explored for signs of diachronic change in the use of three expressions of passive voice in written American English. The articles each have their own focus on specific time periods and genres. In the thesis overall, the corpus material spans nearly a century and a half (1870–2010) and includes five genres of written American English (Fiction, Newspapers, Magazines, Non-fiction Books, and Soap Opera Dialogue), although all five genres are not present in all time periods studied.[1] The empirical methods employed in the thesis are hoped to enrich our understanding of the use of passive voice in American English, to add further nuance to previous theoretical literature and corpus studies, and to offer classification methodologies that may be useful in future studies of naturally-occurring linguistic material.

While each separate article has its own clearly-articulated research questions (see section 5), the overarching aim of the thesis is to describe the use of passive voice in American English diachronically and across genres using corpus methods. Syntactic and semantic claims about the BE-passive, GET-passive, and prepositional passive are investigated empirically in rigorously collected data sets from different time periods and genres. The canonical BE-passive is both considered in its own right and used as a baseline against which the more marked GET- and prepositional passives are compared. Diachronic studies are interpreted in light of established theories that relate to language variation and change, such as prescriptivism, colloquialization, and grammaticalization.

In this introductory survey to the thesis, I take the opportunity to discuss the overall findings of the thesis and to go into a more detailed literature review and background than the article format allows for. This is perhaps especially true of the sections on prescriptivism, colloquialization, and grammaticalization, which are of great relevance to my topic, but which there is limited space to dwell on in shorter articles (sections 2.2 and 6.2). Space is also devoted to a discussion of corpus-linguistic methods (section 3.1) and why it was felt to be appropriate to use these particular language corpora and genres to try to answer my research questions (section 3.2). The corpus methodology called for clearly-defined searches (section 3.3) and the

---

[1] In Article 1 (Schwarz 2015: 156), I also use the spoken-language sections of the BYU version of the British National Corpus (Davies 2004) and the Corpus of Contemporary American English (Davies 2008) for purposes of comparison, but only as a side note, and no data were collected from these spoken-language sources for further analysis in the thesis.

development of rigorous, replicable, classification systems. In section 4, these classification systems are presented and discussed in terms of their possible application in future studies. The four articles are individually summarized in section 5. Overall findings of the thesis, with all of the articles taken together, are presented after the individual summaries, in section 6. A brief, summarizing conclusion with possible directions for future research is found in section 7.

2. Background to the thesis

The use of passive constructions has been undergoing a great deal of change in recent centuries. Developments such as the declining use of the BE-passive and the rise of the GET-passive have been shown to be especially dramatic in American English (see sections 2.1 and 2.2). However, the focus on American English (rather than, say, Australian, Indian, or British English) should not really require explanation. Each variety of English is equally worthy of study in its own right. This is not a cross-variety study, and the results here are not compared to any other variety of English, although this is recognized as a very promising direction for future research.

This section begins with a review of earlier research on passives by way of background. The literature review on passive voice in this section is broken down as follows: a broad definition of passive voice (section 2.1) is followed by separate sections which offer brief accounts of each of the passive expressions considered in this thesis: the BE-passive (section 2.1.1), the GET-passive (section 2.1.2), and the prepositional passive (section 2.1.3). The three major contemporary grammars of English are drawn on heavily in these sections (see Mukherjee 2006 for an insightful overview): Quirk et al.'s 1985 *Comprehensive Grammar of the English Language*; Biber et al.'s 1999 *Longman Grammar of Spoken and Written English*; and Huddleston & Pullum's 2002 *Cambridge Grammar of the English Language*.

In this thesis, thousands of passive sentences have been classified along a variety of syntactic and semantic parameters and meaningfully interpreted according to relevant linguistic theories. In section 2.2, I present three theoretical concepts relevant to changes in the use of passives: prescriptivism, colloquialization, and grammaticalization.

2.1. Passive voice

Quirk et al. (1985: 159) define voice as "a grammatical category which makes it possible to view the action of a sentence in either of two ways, without change in the facts reported." English has two grammatically-

marked voices: active and passive.[2] Assuming that active is the default voice category, a passive sentence is a sentence in which an object noun phrase (NP) has been "promoted" to subject position. Example (1), from the Corpus of Historical American English (COHA) (Davies 2010), is of an active sentence, with the direct object in its "usual" place, and the constructed example in (2) gives the passive version of that sentence, in which the active object is found in subject position. The subject of the active sentence may or may not be present in passive voice; if present, it is found at the end of the sentence in an optional *by*-phrase, as in (2).[3] For ease of reference, verb phrases in example sentences are underlined. Noun phrases of interest, such as subjects, objects, and prepositional complements, are in bold.

(1) Well anyway, the governor <u>nominated</u> **her**. (COHA Fiction 1958)
(2) Well anyway, **she** <u>was nominated</u> (by the governor).

Students of English are usually taught how to form a passive sentence in the following way: they are to place the object at the beginning of the sentence, then use auxiliary BE plus a past participle followed by an optional agent prefaced with *by*, as in sentence (2). The full range of options for English passives is a little wider, and encompasses, besides the passive with BE (*the book was sold*), the GET-passive (*the book got sold*), and the prepositional passive (*the book was sent for*). Each of these kinds of passive is described in greater detail in the separate sections which follow. English also includes a construction with similar object-promoting properties known as the mediopassive, as in *the book is selling well*: this construction is beyond the scope of this thesis, but readers who are interested in finding out more about mediopassives are directed to Hundt (2007) for a helpful guide.

Despite the well-documented "fear and loathing" of the English passive (Pullum 2014; cf. section 2.2.1), from an information-structural standpoint, passive voice enables English speakers (and writers) to convey a useful shift in focus where necessary. Biber et al. (1999: 477) write that "one of the major functions of the passive is that it demotes the agent of the verb (often the person doing the action of the verb), while giving topic status to the affected patient (the entity being acted on)." Passives are thus often used where the agent (the NP that would be the subject of the active version of the sentence) is either uninteresting or very obvious (in which case it is normally omitted); or where it makes sense to place it at the end of the sentence for discursive reasons such as "given vs. new information" (Pullum 2014: 64). Consider

---

[2] While English only has active and passive voice, many other languages also have a grammatically-marked middle voice. See Barber (1975) for an early and useful account of the typology of voice.
[3] Huddleston & Pullum (2002: 1428) have an unusual analysis of the prepositional *by*-phrase as an "internalised complement" in the VP rather than a separate clause element in a passive sentence.

the longer examples in (3) and (4), taken from the TIME Magazine Corpus (Davies 2007):

> (3) Mr. Vauclain is a solemn looking man, tall and well set up. He wears a "cutaway" as a uniform and looks not unlike a bishop. He works from 7 A.M. to closing and can be seen by anybody at any time. Born in Philadelphia in 1856, **he** was educated at the University of Pennsylvania and began work at the Altoona shop of the Pennsylvania Railroad. (TIME 1920s)

> (4) The real question between the nation and the Stuarts was whether the king or the Parliament should rule. **That question** was settled, after a series of convulsions, by the Revolution of 1688. (COHA Magazines 1870s)

In example (3), not only is the existence of *Mr. Vauclain* "given information" by the time we get to the passive *he was educated*, but it is entirely uninteresting to the reader to know who educated him; we assume based on our background knowledge of the world that it was educators at the University of Pennsylvania, and it would be nothing short of weird to have it spelled out for us. Sentence (4), which includes a *by*-agent, also exemplifies the common cohesive technique of "given vs. new." *That question*, which refers back to the preceding sentence, is given information. It would have been slightly more difficult to process the text if the sentence had begun with new information and instead read: *The Revolution of 1688 settled that question after a series of convulsions.* These corpus examples further illustrate what Pullum (2014: 73) has already clearly demonstrated: namely, that "the claims about why you should avoid passives—the allegations about why they are bad—are all bogus, and the interesting point (the discourse condition) is always missed" (on which more shortly in section 2.2.1).

This is not to say that the use of passive voice is always a good idea, as discourse needs must inform syntactic choice. The information-structural properties of the passive can obviously also be used to inappropriately suppress the role of a participant. The psychologist Gerd Bohner (2001) has studied the linguistic topicalization of female rape victims in sentences such as *a woman was attacked* (a sentence which also diminishes the crime of the agent) versus active sentences such as *a man attacked a woman*; but such studies reveal important biases in writer or speaker perspective rather than systemic flaws in English syntax.[4]

---

[4] Bohner (2001), whose main field of study is not linguistics, did not include GET-passives in her study; doing so might have had an even greater influence on her participants' tendency to blame the victim, as the GET-passive is said to impart a sense of subject responsibility (see section 2.1.2).

2.1.1. The BE-passive

The most canonical of the English passives is of course the BE-passive, formed as described above, with auxiliary BE followed by a past participle, as in examples (2)–(4). The BE-passive has been in use for a good deal longer than the GET-passive and even the prepositional passive, and seems to have descended from a similar construction in Old English that could be formed with the auxiliaries BEON and WESAN (Denison 1993: 414–415).[5] The BE-passive also differs greatly from the GET-passive in its association with formal, written language, as clearly shown in Biber et al. (1999: 397). This formality, however, does not seem to be the cause of the vociferous condemnation that the construction has recently fallen under (section 2.2.1).

Despite its information-structural usefulness, the BE-passive has been declining in frequency in written genres over the last century, especially in American English (Mair 2006a: 190; Leech et al. 2009: 153). The BE-passive, as the canonical English passive, is the passive construction that people are most likely to recognize and the one that is most explicitly proscribed. The decrease in the use of the BE-passive in writing is varyingly attributed to prescriptivism (section 2.2.1) and to colloquialization (section 2.2.2). Even though the BE-passive is decreasing in frequency, it is still by far the most common expression of passive voice in English, and is used in a wider variety of syntactic and semantic contexts than GET- and prepositional passives are. For this reason, BE-passives are often used as a control group in the studies in this thesis: claims made about GET- and prepositional passives are checked against data sets of BE-passives, the more established construction, in order to find out whether proposed constraints on these newer passives really are particular to them alone, or are simply true of all passives in general.

2.1.2. The GET-passive

The GET-passive, as in example (5), while much rarer than the BE-passive, has been increasing dramatically in frequency in written English over the last few decades (Mair 2006a: 113, Leech et al. 2009: 156).

> (5) **The girl who** <u>gets rescued</u> from the fire, and who indirectly caused it, is Susan Hayward. (TIME 1940s)

The GET-passive is a relatively recent innovation compared to the much more established BE-passive. The first attested example of a GET-passive that

---

[5] Old English also had a mutative passive that was formed with the auxiliary WEORÐAN ('become'). While this older passive construction bears semantic similarity to the modern-day GET-passive, it is not its predecessor, having fallen out of use centuries before the first attested GET-passive (Denison 1993: 414–415; Mitchell & Robinson 2012: 105).

13

is currently suggested by the *Oxford English Dictionary* (*OED*), shown in (6), is from 1568.[6]

> (6) 1568   T. North tr. A. de Guevara *Dial Princes* (rev. ed.) ii. 114 (*OED* s.v. *get* v.29b)
> If hee bee not his kynseman, or neere allied, let him yet at least get acquaynted with him.

Example (6), however, is not really an example of a passive sentence and must be disqualified as there is no active version of this sentence which is not reflexive (see also Denison 1993: 420; section 4.1 of this survey). The two examples with *get loosed* from 1584 and 1628 which follow are also not examples of true, verbal passives. The first clear example of a GET-passive given in the *OED* is from 1665, and is reproduced in (7).[7]

> (7) 1665   J. Winthrop Let. in H. Oldenburg *Corr.* (1986) XIII. 403 (*OED* s.v. *get* v.29b)
> These are a small black Caterpillar that breed in the very bud of the tree... Some [trees] also get killed by it.

It was earlier thought that passive uses of GET + past participle such as (7) arose from transitive-reflexive uses of GET such as *get oneself dressed/introduced* (Givón & Yang 1994: 139). However, a more recent corpus-based study by Fleisher (2006) has shown that the GET-passive has its origins in GET + adjective constructions like *get sick*, where ambiguous participial adjectives (*get worried*) paved the way for purely verbal participles (*get killed*). The fact that passive GET has its origins in adjectival constructions has implications for the grammaticalization trajectory of the GET-passive, as discussed in section 6.2.

The three major contemporary grammars of English (Quirk et al. 1985; Biber et al. 1999; and Huddleston & Pullum 2002) agree that GET can be used to form true, verbal passives such as (5) and (7). However, as grammaticalized as this use of GET might be, it is still not possible to count GET among the primary auxiliary verbs of English. I have been calling GET an "auxiliary" for the sake of convenience, as do Quirk et al. (1985: 160) and Biber et al. (1999: 475). However, GET is not a true auxiliary verb, as it does

---

[6] Examples (6) and (7) are from the entry: "get, v.29b." OED Online, Oxford University Press, June 2017, www.oed.com/viewdictionaryentry/Entry/77994. Accessed 17 October 2017.
[7] Interestingly, this first clear example of an unambiguously verbal GET-passive in the *OED* is from early American English. John Winthrop the Younger followed his father, the founding governor of the Massachusetts Bay Colony, to the new world at about age 25. His GET-passive occurs in a letter written to Henry Oldenburg of the Royal Society in London, of which Winthrop was a Fellow.

14

not display the "NICE" properties (Negative, Interrogative, Code, Emphasis) and thus does not qualify as an operator (Palmer 1965: 15, 21; Huddleston 1976: 333–334).[8] "Auxiliary" GET cannot precede *not* in negative sentences or precede the subject in interrogative sentences (the N and I of NICE, as shown in (8) and (9)); it cannot be a place holder in cases of ellipsis (the C, or "code" test, example (10)); and it supposedly cannot carry emphasis (the E, example (11)). Instead, dummy operator DO must be used in the NICE functions.

(8) She wasn't selected for promotion.
     *She gotn't selected for promotion.
     She didn't get selected for promotion.

(9) Was she selected for promotion?
     *Got she selected for promotion?
     Did she get selected for promotion?

(10) Was she selected for promotion? She was.
     *Got she selected for promotion? *She got.
     Did she get selected for promotion? She did.

(11) She *was* selected for promotion.
     ?She *got* selected for promotion.[9]
     She *did* get selected for promotion.

The GET-passive is thus a less fully grammaticalized passive auxiliary in English. There are further differences between GET- and BE-passives. One of the most important of these is frequency. Even in the most recent, speech-related material used in this thesis, the BE-passive was about ten times as frequent as the GET-passive (see Article 1: 163). The GET-passive also differs from the BE-passive in that it is an informal construction which is more common in spoken language than in writing. Biber et al. (1999: 476) report that, in their Longman corpus, the GET-passive "occurs only in conversation, except for an occasional example in colloquial fiction." Other corpus investigations reveal that, while GET-passives do occur in written English, they

---

[8] See also Quirk et al. (1985: 120–127) for a characteristically thorough description of operator function.

[9] I have marked emphatic GET (11) as questionably grammatical as I cannot completely ignore my own native-speaker intuition; my feeling is that such emphasis might in fact be possible, as in the following imagined exchange: *Hasn't he been out of work for like five years? Oh, he* got *hired at Safeway, but he only kept that job for a few months.* This would correspond to Quirk et al.'s (1985: 124) "emphatic positive," one of their tests for auxiliary-as-operator and the "E" of the "NICE" test. It is possible that this is a low-status variant, but nonetheless linguistically valid. Furthermore, if I am correct, this would be a sign of further grammaticalization of auxiliary GET as it proceeds along the grammaticalization cline given in section 2.2.3. This is an exciting possible avenue for future research which would probably require additional methods to complement the corpus-linguistic approach.

tend to be more frequent in the genres identified as more "agile" by Hundt & Mair (1999), such as fiction and newspapers (see also Leech et al. 2009: 154–158 and Mair 2006a: 111–117; Articles 1–3 in this thesis).

There have also been special semantic restrictions proposed for GET-passives; for one thing, GET-passives are supposed to confer a sense of adversativity on a situation (Chappell 1980; Quirk et al. 1985: 161; Biber et al. 1999: 481; Carter & McCarthy 1999; Huddleston & Pullum 2002: 1442), which amounts to a sense that whatever is happening to the passive subject is undesirable, as in sentence (7) with *get killed*. However, sentence (5), with *gets rescued*, offers evidence that GET-passives can also be benefactive. Recent studies such as Leech et al. (2009:157) and Article 1 of this thesis support the tendency of GET-passives to have non-neutral semantics in this respect. Importantly, however, there have not to my knowledge been previous studies of the semantics of the BE-passive to compare GET-passives with.

Finally, GET-passives have been said to be agentive (Chappell 1980; Sussex 1982; Collins 1996; Downing 1996; Huddleston & Pullum 2002: 1442). Agentivity means that the passive subject bears some responsibility for whatever is happening to it; in a sentence like *he got shot* we are supposed to infer that *he* did something to provoke the shooting, whereas the version with BE, *he was shot*, would supposedly be more neutral. Agentivity (or subject responsibility, as it is sometimes called) is difficult to search for in a corpus, and might be better investigated using a different methodology; but the occurrence of GET with human subjects was implemented as one means of exploring this constraint in Article 3.

2.1.3. The prepositional passive

English is one of the few languages which allow for the prepositional passive, as in (12), where the object of a preposition is found in subject position rather than following the preposition, as it would be in the active version shown in (13).

(12) Shall **this gramarian** <u>be listened to</u> in his own words? (COHA Fiction 1810, spelling original)
(13) Shall we <u>listen to</u> **this grammarian** in his own words?

The only other languages in which prepositional passives are attested are the Scandinavian languages (excepting Icelandic),[10] Vata and Gbadi (spoken on the Ivory Coast), and some varieties of North American French (Koopman 1984 and Abels 2003 as cited in Findlay 2016: 256). Because prepositional passives are so typologically rare, linguists have been interested in the conditions that allow them to exist in languages at all. The prepositional passive first appears in English documents from the Middle English period, and only

---

[10] See Engdahl & Laanemets 2015 for a corpus study of Scandinavian prepositional passives.

seems to have become widespread in Early Modern English (Seoane Posse 1999; Goh 2001; and Dreschler 2015).[11] While Denison (1993: 143) argues that the loss of case marking in English is *not* behind the innovation of the prepositional passive, he argues that it did, at least to some degree, allow for its propagation (or spread) in the language. Further support for this idea is found in the fact that the one Scandinavian language which does not have prepositional passives, Icelandic, is also the one Scandinavian language which remains comparatively synthetic.

A major issue in analyzing prepositional passives is determining constituency: [verb + preposition] [noun phrase] or [verb] [preposition + noun phrase]. Huddleston & Pullum (2002: 276) argue for these being two wholly discrete categories: Type I prepositional passives (passives in which the verb + preposition form a lexicalized unit); and Type II (passives in which the preposition is selected by the noun phrase). Findlay, on the other hand, (2016: 257) argues (a) that this divide is unsatisfactory in that prepositions have some meaning in and of themselves, and (b) that Huddleston & Pullum's two Types "are two sides of the same coin, rather than totally separate phenomena." Representing a middle ground between Huddleston & Pullum and Findlay, Quirk et al (1985: 1164–1165) and Brinton & Traugott (2005: 128) suggest gradient scales wherein the preposition may be more or less closely associated with the verb.

It is possible that corpora can be of some help in investigating whether or not the verb + preposition in prepositional passives form one lexicalized unit. In his early work on prepositional passives, Bolinger (1975: 59) suggests that the frequency of occurrence of verb + preposition combinations could offer a measure of their degree of lexicalization; a question that is much easier to investigate in modern-day computerized corpora than it would have been at the time he suggested it. However, Bolinger (1975: 58–59) also argues that the status of the verb + preposition as a lexicalized unit should not be the focus of too much attention, and that other factors may be of importance in predicting the acceptability of prepositional passives.

Specifically, Bolinger (1975) argues that certain discourse characteristics and semantic features promote the acceptability of prepositional passives. The discourse features that were supposed to enhance the interpretability and thus acceptability of prepositional passives include coordination with another passive verb, and [light verb] + [noun] + [preposition] combinations such as *made use of*, where the light verb is supposed to lead to an expectation that a noun will follow.

---

[11] Seoane Posse (1999) and Dreschler (2015) both use quantitative corpus methods to trace the spread of the prepositional passive in earlier English. These are the only prior corpus studies (in the modern sense of the term) of the prepositional passive in English that I am aware of. Couper-Kuhlen's (1979) study is not a corpus study in the modern sense of the term, but offers admirable insights and empirical methodology nonetheless.

The main semantic feature that Bolinger treats is "affectedness." Bolinger writes that a prepositional passive sentence is possible when "the subject […] is conceived to be a true patient, i.e., to be genuinely affected by the action of the verb." This assertion is echoed in Quirk et al (1985: 1164–1165), who write that, rather than proving that the verb + preposition form a lexicalized unit, "the passive is primarily an indicator of the fact that the prepositional complement is being treated as an *affected* participant in the clause." This affectedness requirement is supposed to explain why example (14) is felicitous, but example (15) is not (both from Quirk et al. 1985: 1165).

(14) The problem was gone into.
(15) *The tunnel was gone into.

Attempts have been made to describe how one might test for affectedness (see Beavers 2011), but these tests are difficult to apply to naturally-occurring language data; this problem is addressed at length in Article 4.

Examples (14) and (15) also highlight the fact that prepositional passives tend to have abstract rather than concrete subject NPs. This idea was proposed by Svartvik (1966: 165) and is found in Quirk et al. (1985: 163), and, in my Article 4, turned out to be a major hindrance in applying affectedness tests of the kind offered in Beavers (2011).

Huddleston & Pullum (2002: 1446) do not discuss affectedness as a feature of prepositional passives. Rather, they write that their Type II prepositional passives (those where the verb + preposition do not form a lexicalized unit) "are felicitous only if the VP indicates either a significant property or a change in a significant property of the subject-referent." They illustrate their first point, about the "significant property" of the subject-referent, with the following sentence:

(16) This bed was slept in by George Washington.

In (16), "the fact that George Washington slept in the bed gives it some historical interest," according to Huddleston & Pullum (2002: 1446). Davison (1980: 54) predates Huddleston & Pullum with this idea, writing that there is a tendency in prepositional passives for the subject-referent to have "some perceptible property" and "if a famous person is mentioned as the agent, the passive sentence conveys the suggestion that the subject-topic has the quality of being interesting, at least to the speaker, by virtue of its connection with this person." The idea of perceptual salience of the subject-referent seems thus a possible factor in the acceptability of prepositional passives.

The stylistic acceptability of the prepositional passive may also be possible to investigate in corpora; in this thesis, the diachronic frequency of the prepositional passive compared to the non-prepositional passive was of in-

18

terest in researching the role of prescriptivism in language change in Article 4. The prepositional passive, like the GET-passive, is potentially doubly sensitive to prescriptivism. Not only is it a passive, which is a much-hated construction in the U.S. (see Pullum 2010, 2014; section 2.2.1); it also involves so called "preposition stranding," where a preposition follows rather than precedes its NP object (see Yáñez-Bouza 2015).

2.2. Diachronic development: Prescriptivism, colloquialization, and grammaticalization

In this thesis, diachronic developments in the use of the BE-passive, GET-passive, and prepositional passive are interpreted against the theoretical backgrounds of prescriptivism, colloquialization, and grammaticalization. Corpus exploration has proven to be a good method for empirically investigating these three concepts as they relate to the diachronic development of language (see Hopper & Traugott 2003: 36; Leech 2004: 73; Mair 2006a; Mair 2011; Anderwald 2016). While the published, written language found in the corpus material used in this thesis may be seen as less "natural" than casual conversation, written material is of particular usefulness here.[12] Prescriptivism and colloquialization are especially (in the case of colloquialization, exclusively) relevant to written language; also, one means of exploring both colloquialization and grammaticalization is by tracing frequencies of a linguistic feature across styles and genres (Hopper & Traugott 2003: 232).

Prescriptivism, colloquialization, and grammaticalization are topics of current interest in corpus linguistics, and quite a lot of recent work has gone into defining and disentangling these processes. A brief introduction to each is given here; these are by no means all-encompassing, and our understanding of these three topics is continuing to evolve. In section 6.2, I return to these three concepts as they relate to the findings in this thesis, and propose ways in which the results of my studies enhance our understanding of how they may work together to either promote, slow, or reflect linguistic change.

2.2.1. Prescriptivism

Prescriptivism highlights a divide, whether real or imaginary, between linguistic scholars and the so-called "general public." Broadly speaking, linguistic prescriptivism constitutes an effort by supposed experts (whether language pundits, writers of style guides, newspaper editors, teachers, activists, or the grammar police of internet memes) to control language use. Descriptive linguistics, on the other hand, which is what linguistic scholars aim to engage in, entails a more neutral approach to language, with a focus on investigating actual use rather than dictating "appropriate" usage rules and

---

[12] Mair (2006a: 183) calls informal face-to-face conversation the "'natural' linguistic baseline." I will return to this point in section 3.1.

norms. For a linguistic scholar, then, there is no "bad" language. For a prescriptivist, there most certainly is, and passive voice is one of the baddest objects of their recent disdain.

Cameron (1995) problematizes the supposed divide between prescriptivists and descriptivists and suggests that linguists re-examine their anti-prescriptivist stance. She argues (a) that non-experts have every right to be part of the discourse about language use and (b) that linguists need to recognize that they, too, have engaged in normative activities such as language planning, thereby influencing language use. "On that level," she writes, "'description' and 'prescription' turn out to be aspects of a single (and normative) activity: a struggle to control language by defining its nature" (Cameron 1995: 8).

Lippi-Green (1997: 8–9) takes issue with Cameron's stance, arguing that here, Cameron is overlooking a key difference between descriptivism and prescriptivism:

> There is a qualitative difference between the two approaches […]. The linguist and the non-linguist claim different kinds and sources of authority to validate their individual approaches to language. Linguists are often impatient when they are cornered at cocktail parties and asked to debate language issues which to them brook no debate, just as geologists and biologists would be hard-pressed to debate (with any degree of seriousness or interest) arguments against evolutionary theory based on the writings of the Bible. Linguists claim some authority in the description of language based on observation, experimentation, and deduction, so that the claim *All living languages change* is not a matter of faith or opinion or aesthetics, but observable fact (which is not to say that all claims by linguists are equally supportable by fact).

Whether prescriptive efforts to control language must always be viewed as negative, however, is a question which deserves some discussion here. In a more recent work, Curzan (2014) nuances our definition of prescriptivism and explores its continuing manifestations. She distinguishes four "strands" of prescriptivism, reproduced below (Curzan 2014: 24):

- Standardizing prescriptivism: rules/judgments that aim to promote and enforce standardization and "standard" usage.
- Stylistic prescriptivism: rules/judgments that aim to differentiate among (often fine) points of style within standard usage.
- Restorative prescriptivism: rules/judgments that aim to restore earlier, but now relatively obsolete, usage and/or turn to older forms to purify usage.
- Politically responsive prescriptivism: rules/judgments that aim to promote inclusive, nondiscriminatory, politically correct, and/or politically expedient usage.

The fourth strand, "politically responsive prescriptivism," stands out from the others because such attempts to control language use, such as requiring government officials to use gender-neutral language, promote inclusiveness and equality. However, the proscription on the use of passive voice does *not* fall under this inclusive strand, but probably under the second strand, "stylistic prescriptivism." The anti-passive sentiment has recently grown so fervent that the Microsoft grammar checker actually counts it as a style "error" (Curzan 2014: 76–77, 81). This very extreme instance of stylistic proscription is clearly neither inclusive nor evidence-based.

Pullum (2010; 2014) has explored, at impressive length, the vociferous and unfounded distaste that style manuals have for passive voice, and has shown (a) that this distaste is near ubiquitous and (b) that "passive voice" means something different to linguistic scholars than to basically everyone else. Even self-styled "grammarians" are frequently unable to produce an example of a passive sentence, and Pullum's (2014: 67) extravagant lament that "the general public's education regarding the notion 'passive voice' is nothing short of disastrous" seems accurate, if dramatic.

It is the BE-passive (whose decline was evident in all four of the articles in this thesis) that tends to be singled out by prescriptivists as the one and only linguistic expression of passive voice in English (Pullum 2014: 62). The GET-passive (Articles 1–3) and the prepositional passive (Article 4), while less well-known outside of linguistic research, are also occasionally implicated under other proscriptions: the relatively common proscription against the use of the word GET in formal writing (documented thoroughly in Anderwald 2016: 227–234), and the longstanding proscription against the so-called "stranding" of a preposition (see Yáñez-Bouza 2015).

The most well-known American style manual of all, the infamous yet popular *Elements of Style* (Strunk & White 2000: 77–78), still inexplicably includes its original proscription on preposition stranding under the section "Avoid fancy words." Their advice on preposition stranding is, according to Pullum (2010: 40), "not just atavistic but flagrantly inaccurate." They also still include a one-sentence proscription on the form *gotten* (Strunk & White 2000: 48), though not on GET-passives as such. GET-passives can otherwise be seen as doubly proscribed (avoid passive voice + avoid GET); however, this assumes not only that writers are able to correctly identify passives (which Pullum 2014 rightly questions) but that writers have been responding to the explicit proscriptions on GET. The GET-passive has rarely been referred to by name in style guides, but the use of the word GET (especially in its more functional senses) has indeed been widely proscribed in American grammars (Anderwald 2016: 227–234). Anderwald (2016: 245) cautiously suggests that the proscription on GET may have been successful in slowing down the tremendous increase in the use of GET, but her own COHA data (2016: 220) force her conclusion to be a tentative one.

The reasons for proscribing GET (and, by association, the GET-passive) are shown by Anderwald (2016: 233–234) to stem first and foremost from an objection to its polysemous nature as a grammaticalizing item (on which more shortly). That is, the newer function-word meaning is not the "real" one, and GET should only be used in its original sense of 'obtain.' The other major objection in style guides to the word GET is that it is too "vulgar" and "colloquial," which, while undeservedly judgmental, does show a recognition of functional GET's origins in spoken language (see section 2.1.2).

Explicit proscriptions of passive voice are not only often inaccurate; they are, as I have hinted at, also rather oddly motivated. As stated above, the "passive voice" referred to in style manuals (where identified correctly) tends to be the BE-passive. Interestingly, the BE-passive is known to descriptive linguists as being typical of formal, written English. However, style manuals do not seem to object to the construction's formality level, but rather to its perceived "lack of clarity" and "wordiness." The BE-passive is blamed for making a piece of writing "weak, dull, vague, cowardly, bureaucratic, and dishonest" (Pullum 2014: 60). As recently as August 2017, *The New York Times* ran an opinion piece called "Tips for aspiring op-ed writers" (Stephens 2017) that contained the following item on a numbered list of "tips":[13]

> Avoid the passive voice. Write declarative sentences. Delete useless or weasel words such as "apparently," "understandable" or "indeed." Project a tone of confidence, which is the middle course between diffidence and bombast.

While the first sentence of Stephens' "tip" would seem to be about the use of passives, none of the sentences that follow in this item appear to have anything to do with voice, but rather with the kinds of "weakness" and "wordiness" that Pullum (2014) has identified as the typical style-guide objections to passive voice. The fact that the proscription on the passive never directly seems to have to do with the snobbery/formality of the construction has implications for whether prescriptivism or colloquialization can be said to be responsible for the diachronic decline in BE-passives; this point will be returned to in section 6.2.

Whether prescriptivist attitudes are merely a reaction to ongoing language change or actually have the ability to influence the use of certain linguistic features is a matter of some debate (see Curzan 2014 for a thorough discussion). While Leech et al. (2009: 151) reasonably suggest that prescriptive advice might be playing a role in the decline of the passive,[14] there are rea-

---

[13] Stephens (2017) actually does manage complete avoidance of the passive voice in his column. However, the excessive use of imperatives he engages in to compensate for this, while appropriate for the instructive nature of a list of tips, can hardly be seen as a model for aspiring op-ed writers.

[14] I make the same tentative suggestion in the articles included in this thesis, but still feel it needs further investigation.

22

sons to question the efficacy of the advice when the average American is unable to recognize a passive sentence (Pullum 2014: 67). Furthermore, Anderwald (2016: 245), who also uses corpus methodology, finds little evidence that nineteenth-century grammar writing actually influenced usage, but suggests that the GET-passive may be an exception, and that the proscription has slowed its rise. This claim is impossible to definitively test in a corpus, as we cannot know what the development of the GET-passive would have looked like if there had been no proscription against it. In section 6.2, I will reflect on what, if any, evidence there is in the four studies in this thesis for whether or not prescriptivism is having any effect on the use of the passive.

2.2.2. Colloquialization

The decline in the BE-passive seen in all four of the studies in this thesis is varyingly attributed to prescriptive influence (Leech et al. 2009: 151) and to colloquialization (Leech 2004: 73; Mair 2006a: 190; Leech et al. 2009: 244), and there is no real reason why it should not have to do with both of these. The term colloquialization to describe a stylistic shift in written language towards spoken usage is found first in Siemund (1995: 357) and Mair & Hundt (1995: 118), and was later popularized by Mair (1997: 203–205). However, an even earlier description of the process may be found in Fairclough (1992: 204), who writes of a "shift […] towards speechlike forms in writing"; this is essentially a brief definition of what we now call colloquialization. It has more recently been neatly defined by Farrelly & Seoane (2012: 394) as "the tendency for written language to incorporate features of the spoken language," which accords nicely with Fairclough's earlier idea. Colloquialization is thus not a force for linguistic change, but a process that written language may undergo. As such, it is closely related to the process of genre drift (Biber & Finegan 1989). The actual mechanism of change that leads to colloquialization must be democratization.

Fairclough (1992: 201) defines democratization in discourse as "the removal of inequalities and asymmetries in the discursive and linguistic rights, obligations and prestige of groups of people." He goes on to offer five areas of discursive democratization, which I reproduce here:

- Relations between languages and social dialects.
- Access to prestigious discourse types.
- Elimination of overt power markers in institutional discourse types with unequal power relations.
- A tendency towards informality of language.
- Changes in gender-related practices in language.

Fairclough's fourth area of discursive democratization, "a tendency towards informality of language," has obvious relevance for colloquialization.[15] Mair (1997: 203), Leech et al. (2009: 259) and Farrelly & Seoane (2012) have all problematized the relationship between democratization, colloquialization and informalization. I would like to take this opportunity to reflect on the emerging distinction between these concepts as they relate to the use of passive voice in American English, as relevant to the research questions posed in Articles 1–3.

Leech et al. (2009: 258) and Farrelly & Seoane (2012: 393) have a much narrower definition of democratization than I have given above, only counting Fairclough's third point about the "elimination of overt power markers" as "democratization proper." They may have based their definition on a shorter explanation of the democratization of discourse that comes earlier in Fairclough (1992: 98). As he specifically refers the reader to the later chapter, I believe that he intends the full list of five "areas" to be included as manifestations of discursive democratization. Like Farrelly & Seoane (2012: 393), I consider colloquialization and informalization to be subsumed under democratization; but I believe what they call "democratization proper" is actually closer to the kind of deliberate, prescriptivist efforts to promote inclusive language described by Cameron (1995) and Curzan (2014) (as in examples of gender-inclusive speech from official agencies, etc.). Therefore, democratization subsumes colloquialization, informalization, and *certain, inclusive, kinds of prescriptivism*. However, these are *not* the kinds of prescriptivism that relate to the use of passive voice, as it is not the formal nature of the construction that tends to be stigmatized; this is a point I will return to in section 6.2.

While Mair (1997: 203–204) seems to regard informalization as a greater societal trend which has a particular manifestation in written language in the form of colloquialization, Leech et al. (2009: 239f.) and Farrelly & Seoane (2012: 395) seem to differentiate colloquialization and informalization as two different trends that both apply solely to written language. Colloquialization relates specifically to the adoption of spoken-language features, whereas informalization relates to the spread of informal features that do *not* have their origins in spoken language.[16] By this definition, the decrease in BE-passives (Articles 1–4) and the increase in GET-passives (Articles 1–3) are both clearly candidates for signaling colloquialization rather than informalization, as written language shifts towards speech in both cases.[17]

---

[15] The third and fifth points also have special relevance to prescriptivism.

[16] Farrelly & Seoane (2012: 395–396) also seem to take the view that certain text-types undergo colloquialization while others undergo informalization; this line of inquiry is not pursued in this thesis.

[17] Leech (2004: 73) and Leech et al. (2009: 244) call the decline in BE-passives a "negative manifestation of colloquialization."

24

Furthermore, capitalism (which is not necessarily a result of democratization) would also seem to be driving colloquialization. This is probably especially true of American English, which is the variety in focus in all of my studies. I say that capitalism plays a role because colloquialization is especially likely to occur in more "agile" written genres (as described by Hundt & Mair 1999), i.e. those genres whose readership stands to increase if writers keep their language accessible. Put bluntly, colloquialization sells. Newspaper language is a typically "agile" genre prone to colloquialization. Formal academic writing, on the other hand, constitutes an example of a more "uptight" genre, one which would be likely to resist the adoption of spoken-language features longer, until their presence has become routine in less formal genres; this shades into grammaticalization (see Hopper & Traugott 2003: 232).

2.2.3. Grammaticalization

I am using the term grammaticalization in this study to refer to a language phenomenon rather than a research framework (Hopper & Traugott 2003: 1–2). The language phenomenon called grammaticalization can be broadly defined as the stepwise transition of a lexical item along a cline from lexical towards functional. One classic example of grammaticalization in English is the *going to*-future: *going to* evolves from its lexical motion sense (example (17)) to a grammatical marker of future time (example (18)), undergoing phonetic reduction along the way (the pronunciation *gonna*, which is only possible with the grammatical marker and not the full, lexical motion sense exemplified in (17)).

(17) I'm going to the store.
      *I'm gonna the store.

(18) I'm going to call my grandmother tomorrow.
      I'm gonna call my grandmother tomorrow.

The earlier, lexical 'motion' sense of *going to* is not lost, but coexists in the language alongside the new, grammaticalized sense in a situation known as "layering" (Hopper & Traugott 2003: 49). This layering effect allows for synchronic study of grammaticalization as well. However, in this thesis, only the possible further evolution of GET as a passive auxiliary is explored, and only diachronically. Other uses of GET, as for example in its meaning 'obtain' (sentence (19)) or in its transitive-reflexive use (sentence (20)), are not examined in this thesis. As I discussed briefly in section 2.1.2, Fleisher (2006) has already convincingly demonstrated that the passive auxiliary use of GET derives from its copula + adjective sense, shown in (21).[18]

---

[18] GET is a word with many senses: see the *OED* entry for a more comprehensive list.

25

(19) 'obtain': There was an auction in the village to-day, as I passed through, and I stopped at a cake-stand to <u>get a glass of water</u>, as it was very hot. (COHA Fiction 1872)

(20) transitive-reflexive: "Andy'll <u>git hisself killed</u>!" whimpered Pendrilla. (COHA Fiction 1908)

(21) copula + adjective: The farmer loses the milk; the cow is unhappy with it, and may <u>get sick</u> from being out in the cold storm. (COHA Non-fiction 1915)

While it is interesting to investigate the development of passive "auxiliary" GET versus other uses of GET, as in Hundt (2001), it is also perfectly possible to examine the trajectory of this one, grammaticalizing, sense of GET. Corpora can be searched for signs that a construction such as the GET-passive may be advancing through Hopper & Traugott's (2003: 7) "cline of grammaticality," or the rough steps an item goes through, unidirectionally, on the grammaticalization path:

content item > grammatical word > clitic > inflectional affix

Along this path, the grammaticalizing element goes through a number of well-documented processes. Among these are the kind of phonetic reduction shown in *gonna*, which is not (yet) present for auxiliary GET; and so-called "bleaching," whereby the full semantics of a content word are weakened as it becomes more grammatical. Also, one can look for signs of generalization of grammatical function. Signs of semantic bleaching and morpho-syntactic generalization of the GET-passive were searched for in Articles 1–3 of this thesis.

There is always a danger that a unidirectional cline suggests a very fixed progression with a particular goal as its endpoint. However, Hopper & Traugott (2003: 131) caution against this extreme interpretation of the cline:

> A particular grammaticalization process may be, and often is, arrested before it is fully "implemented," and the "outcome" of grammaticalization is quite often a ragged and incomplete subsystem that is not evidently moving in some identifiable direction.

It is therefore perfectly possible that the GET-passive will be forever "arrested" as a "grammatical word" at the second stage on the cline. But changes along the cline are gradient and not abrupt (see Mair 2004: 121), and there still may be many, smaller changes implicit in that stage; these gradient steps are what was investigated in terms of adversativity, agentivity, and morpho-syntactic generalization in Articles 1–3 in this thesis.

The GET-passive is thus a linguistic feature which is undergoing *further* grammaticalization in its auxiliary-like function in passive sentences (as also noted by Leech et al. 2009: 146);[19] and it is also a feature which is supposed to signal colloquialization in a written genre (see section 2.2.2). These are not conflicting claims. There is, in fact, overlap between the processes of colloquialization and grammaticalization. Increasing frequencies are typical of both processes. There is a tangle here: increased frequency of a linguistic item is both a prerequisite for and a result of grammaticalization (Hundt 2001: 56–57), and, depending on the feature, may also be a sign that a written genre is undergoing colloquialization. Since grammaticalization starts in spoken language, it makes sense that features that are grammaticalizing would also be spoken-language features that might become increasingly acceptable in written genres as they become less marked, leading to overlap of the linguistic features across these two categories. Mair & Hundt (1995: 118), for example, cite the "growing acceptance" of contracted forms and the *going to*-future as examples of colloquialization; these are also commonly regarded as grammaticalizing linguistic items (as in Mair 2011: 245, where he discusses the canonical *going to*-future example). Both processes have to do with linguistic features typical of spoken language; the most important difference is that colloquialization is change in a written text-type or genre (signaled by linguistic features), while grammaticalization is change in a linguistic feature (which can been partly traced by its spread through different genres as discussed in Hopper & Traugott 2003: 232).

3. Method and materials

3.1. Corpus linguistics

In this thesis, research questions were investigated in large, computerized collections of machine-readable texts. All of the texts consisted of written American English of varying degrees of formality (ranging from non-fiction books to soap opera transcripts). My research questions were well-suited to investigation using corpus methodology, and the corpora were selected and explored in a principled, replicable manner. However, the corpus-based approach to linguistic research calls for some discussion here, and (as with any methodology) its advantages and limitations must be carefully weighed.

---

[19] Anderwald (2016: 219) writes that GET is already grammaticalized as a passive auxiliary. I take her point that GET would seem to be able to function in central passives now and to be fully "bleached," in that context, of the original sense from which it is derived and thus have grammaticalized according to the basic definition set forth by Hopper & Traugott (2003: 4). However, not only does "auxiliary" GET not have operator status (section 2.1.2), but the studies in this thesis also suggest that the GET-passive is still taking small steps towards use in wider syntactic and semantic contexts. The evidence for further grammaticalization of the GET-passive in Articles 2 and 3 of this thesis will be addressed in section 6.2.

Even the most speech-like corpus material in my studies, taken from the Corpus of American Soap Operas (Davies 2012), is still written, which means that it is different from spontaneous spoken conversation in some very important ways. In linguistics, casual, face-to-face conversation between equals is considered the most natural form of language (Mair 2006a: 183) for good reasons. Written language is produced under very different circumstances than casual speech is and can be planned and revised in ways that conversation cannot. And possibly the most basic distinction between spoken and written language is that all humans (with brains of average function and assuming adequate input) learn to speak their mother tongues without any conscious effort, whereas reading and writing are language skills that must be explicitly taught. There are also many human languages which *have* no written form. Truly, face-to-face conversation must, as Mair (2006a: 183) points out, be regarded as "the most basic manifestation of language."

Casual conversation is not impossible to study in language corpora. There are, in fact, carefully compiled corpora of spontaneous conversation. Examples include parts of the London-Lund Corpus of Spoken English[20] and of the ICE (International Corpus of English) corpora.[21] However, transcribing and annotating casual conversation is extremely painstaking, and spoken-language corpora tend to be quite small in comparison to corpora containing written texts, which are much easier to digitize. In studies of older material, it is obviously difficult and often impossible to consult empirical spoken-language data. Historical corpus linguists, in an effort to get as close as they can to the casual conversation of their time period, have carefully compiled corpora of speech-related material. Culpeper & Kytö (2010: 17) offer an illuminating discussion of the different kinds of speech-related texts that are included in historical, speech-related corpora such as the Corpus of English Dialogues 1560–1760 (CED).

However, corpora of written language certainly have their uses as well, and there are at least three reasons why my particular research questions were well-suited to empirical investigation in large collections of written texts.[22] First and most importantly, the GET-passive and prepositional passive are so rare that it might have been difficult to find large enough numbers of them to enable quantitative study if spoken-language corpora had been used. (However, there would be obvious value in examining GET-passives in corpora of conversation: a potentially rewarding avenue for further study.)

---

[20] Full references to corpora are found in the reference list at the end of this survey.

[21] The London-Lund corpus contains approximately 500,000 words of spoken English, around 315,000 of which are from conversation, and the conversation section of ICE-GB (Great Britain) contains approximately 180,000 words; intensely valuable material, but very small in comparison to COHA, which contains over 400 million (much more easily collected) words.

[22] I do make use of speech-related written material in Article 1: the Corpus of American Soap Operas (Davies 2012), and there are speech-related aspects of the fiction genre (invented dialogue) and of newspapers and magazines (quoted speech), as well.

The research questions relating to prescriptivism, colloquialization, and grammaticalization lend themselves especially well to corpus-based research. Although prescriptivists are clearly full of advice on both spoken and written language (see, for example, Wolf 2015 on the much-hated "vocal fry" phenomenon, a common proscription which also has misogynistic implications),[23] their advice on passives seems to apply *solely* to writing. I cannot find a single example of advice proscribing passive voice in spoken language (possibly because it is rare in speech, cf. section 2.1.1), but, as shown in section 2.2.1, American style manuals have been vehemently opposed to both BE- and GET-passives in writing for decades.

Colloquialization, as mentioned above, is a theory which pertains only to written language. While using corpora of conversational interactions might help confirm whether a colloquial form (such as the GET-passive) really has its origins in spoken language, any study of colloquialization must reasonably include an investigation of written material from carefully selected genres.

Similarly, the grammaticalization of the GET-passive could be studied in spoken language material. However, this would require a long diachronic span, which is impossible where casual conversation is concerned. Even from a century ago, recordings would be too sparse to allow for generalizations to be made. Furthermore, the very valuable speech-related corpora of the CED type would probably be of limited use in investigating such a rare linguistic feature. In fact, corpora of written language have shown themselves to be especially well-suited to studies of grammaticalization (see Hopper & Traugott 2003: 232; Mair 2004; and Mair 2011). Mair (2004: 121) offers four points of common ground between corpus linguistics and grammaticalization studies. His bulleted list is well worth reproducing here:

- Both approaches give priority to the study of utterances in their discourse contexts rather than abstract systems of underlying rules.
- Both emphasize the importance of frequency data and statistics.
- Both agree that transitions between grammatical categories are gradient rather than abrupt, and that grammatical form and meaning are interdependent rather than constituting separate and autonomous domains.
- Both, finally, became "hot" in linguistics again in the late nineteen seventies and early nineteen eighties after decades of relative neglect.

All four points are both accurate and interesting, but the first three are of most relevance here, and in fact lead nicely into further discussion of the

---

[23] The sound popularly known as "vocal fry" is created when the glottal folds are held loosely together, resulting in a bubbling sound. This glottal state is also known as "creaky voice," which, while not phonemic in English, is phonemic in many Nilotic languages (Ladefoged 1971). Kim Kardashian and Katy Perry are examples of prominent women who have both been publicly maligned for their tendency to speak with vocal fry.

advantages and drawbacks of corpus linguistic methodology. Traditional descriptive linguistics relies heavily on native speaker intuition and invented examples to inform theories about what is possible in a language. The great appeal of corpus-based methodology is that our intuitive ideas can be tested on actual language – language produced for communicative purposes rather than to prove a linguistic point.

The importance of frequency data and statistics is also appealing from a scientific-method standpoint. How can we claim that language has changed without evidence to back up our claim? And after material has been rigorously collected, examined, and classified into patterns, how can we tell whether the trends we are seeing are significant if we do not analyze them statistically? Corpus-based methodology, while neither definitive nor perfect, does allow for more confident conclusions about language change to be drawn. Also, careful description of search and classification methods allows for studies to be replicated by other researchers.

The third point in Mair's list, about the gradience of grammatical categories, is a very important one (see also Aarts 2007: 97 on subsective gradience). Studies on actual rather than idealized language expose the resistance of naturally-occurring language to strict categorization (as shown in Denison 1999), and also help us to track the kind of diachronic change that signifies grammaticalization.

The advantages of using corpus methods to study passives in this thesis were thus that (a) the methodology fits the research aim of investigating the development of passives diachronically in terms of prescriptivism, colloquialization, and grammaticalization; (b) there is sufficient data to allow for investigation of rare forms like the GET-passive and prepositional passive (see also section 3.2); and (c) the naturally-occurring data allow for more reliable conclusions to be drawn.[24] These are all legitimate reasons for me to have used the corpora which will be described in section 3.2. However, corpus methods are not problem-free, and the remainder of this section deals with potential limitations inherent in corpus studies of written language, and an explanation of how these were mitigated where possible.

An ideal corpus is supposed to be balanced, comparable, and representative, but these three desiderata are not necessarily easy to achieve. Representativeness is especially difficult, and the most serious problem for any study which makes use of written corpus material is, as I see it, the problem of representing a population of speakers. Francis (1979: 110), who was

---

[24] It is recognized that "naturally-occurring" is a blanket term that does not reflect the important distinctions between invented dialogue of the type found in the Soap Opera Corpus and expository prose of the type found in non-fiction books in COHA. The distinction drawn here is between language that was invented by linguists in order to prove a linguistic point, and language that was produced for other, communicative purposes. It is the latter that is found in the BYU corpora. See section 3.2 for a description of the differences between the genres included in the studies.

largely responsible for the design of the Brown corpus,[25] offered an early, strikingly optimistic definition of what a linguistic corpus is:

> A collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis.

The quest for representativeness (on which see, for example, Biber 1993a, 1993b) in language corpora is worthwhile, but not expected to ultimately be fruitful, leading to Leech's (2007) calling representativeness a "holy grail" of corpus linguistics. As Leech (2007: 135) points out, "without representativeness, whatever is found to be true of a corpus, is simply true of that corpus – and cannot be extended to anything else." A logical follow-up question is: how can *any* corpus be representative of a population of speakers? In my studies, the attempt made to mitigate the unrepresentativeness of the content was to include several written genres (which also facilitates tracking colloquialization and grammaticalization as features spread through genres of varying formality).[26] However, while this cross-genre perspective makes results a little more reliable, it certainly cannot be said to make my findings representative of all American English speakers (or writers), as only a privileged few are represented in the published, written material in the corpora, however vast the amount of material may be. Even in the most carefully collected and "balanced" corpora, such as the "Brown family," the material has not been claimed to be representative of all speakers of American or British English.[27]

A major problem for all existing language corpora, even corpora which include material from recorded conversations, is that low-status language tends to be underrepresented (Lindquist 2009: 43). Leech (2007) discusses this problem at length, and offers reasonable methods of dealing with it, such as comparing results with other genres and corpora, as I have done. One of these methods is the recognition of the reader of the text as a participant in "an initiator-text-receiver nexus," which Leech calls an "Atomic Communicative Event," or "ACE." Leech (2007: 138–139) proposes including higher proportions of texts that reach wider audiences, which would necessarily mean a more prominent representation of "low-culture" sources, such as

---

[25] See section 3.2 for a fuller description of the Brown family of corpora.

[26] Articles 1 and 2 only made primary use of one genre in their main studies; but other corpora were always consulted so results could be compared.

[27] The original diachronic/cross-varietal Brown quartet based on Brown and LOB consisted of Brown, LOB, Frown, and FLOB, which together offered two synchronic snapshots of two dialects. Earlier and later snapshots are currently being added to enable more confident diachronic conclusions to be drawn (see Hundt & Leech 2016). These corpora, which are highly comparable with one another, include more genres than the BYU corpora and are more carefully sampled. Even though they are not representative of an entire population of speakers, they may well have a higher degree of representativity than the BYU corpora. See Leech (2007) for a more thorough discussion.

pulp fiction. To some extent, the Corpus of American Soap Opera Dialogue and the TIME Magazine Corpus, used in Articles 1 and 2 of this thesis, must be seen as such sources. Furthermore, it is likely that colloquialization will have progressed further in material that is aimed at wider audiences, which encourages the researcher to contrast these sources with other, more formal genres which are intended for a narrower audience.

The inclusion of "low-culture" sources is intended to improve the representativeness of corpora by giving greater weight to texts that are read by more people. However, this still only represents the written production of relatively few speakers. A major problem that remains, even with ACEs taken into account in corpus compilation, is that all written, published language is high status in comparison to casual conversation. Written language has been planned, edited, and sometimes even published; and, even if we choose to recognize the role of the recipient in the ACE nexus, written language is only available to literate speakers (unless the text is read aloud, as is the case with the Soap Opera Corpus). No matter the genre, corpora of written English cannot be said to be fully representative of a speaker population. Empirical methodology notwithstanding, findings in this thesis are interpreted with a large grain of salt, and results should not be understood as generalizable to an entire population.

3.2. The BYU corpora

This thesis makes exclusive use of corpora compiled by Mark Davies and his team at Brigham Young University. I have touched on the reasons for choosing the BYU corpora above, but in this section I offer a more comprehensive description of the BYU corpora, situating them in the world of language corpora by comparing them to other, earlier corpora, and explaining in more detail why they were used in my studies.

There are three main types of language corpora (McEnery & Hardie 2012: 6–13): opportunistic corpora, monitor corpora, and sample corpora. Opportunistic corpora are simply databases of language that have been compiled for purposes other than linguistic research (such as transcriptions of news broadcasts or newspaper archives), but which are made use of by linguistic researchers.[28] The other two types of corpora, namely monitor corpora and sample corpora, are compiled by linguists. The difference between monitor corpora and sample corpora is the way in which the texts are selected. Monitor corpora, such as the Bank of English (BoE) collected by the University of Birmingham, contain large amounts of fairly continuous diachronic

---

[28] An example of a linguist using an opportunistic corpus is Granath (2009), where a CD-ROM of American Broadcast News is used to investigate the use of "good question" as a discourse marker.

data, and are still being added to.[29] Monitor corpora tend to be quite large and can thus be used to search for less common words and structures.

Sample corpora, sometimes called balanced corpora, are compiled according to more rigid sampling principles in an effort to represent a certain language or population of language users (as problematized at length in section 3.1). Criteria such as dialect, genre, and time period are thus all very important when compiling a sample corpus. The Brown family of corpora, all compiled according to the same principles, offer a high degree of comparability between corpora. Because the two 1961 corpora, namely Brown, which consists of American English, and the Lancaster-Oslo/Bergen Corpus (LOB), which consists of British English, include roughly the same number of words from the same 15 text types, any differences between the language in the two corpora offer good evidence for variation between these two standard varieties. The Frown and FLOB corpora, compiled at the University of Freiburg under the guidance of Christian Mair, were sampled according to the same principles but include data from 1991/1992, and thus offer good evidence of diachronic development for the two dialects.

The International Corpus of English (ICE) is another good example of a balanced corpus. The twenty different varieties of English included in ICE are sampled according to the same guidelines, which allows for reliable comparison between them. ICE has the further advantage of being composed of 60% spoken language data. The subcorpora are quite small, however, at only 1 million words apiece, so they are only useful when looking for rather frequent lexemes (see Nilsson [Schwarz] 2013).

At the time of writing, BYU offers free digital access to thirteen different English-language corpora, one corpus of Spanish, and one of Portuguese. Only three of the corpora – the Corpus of American Soap Operas, the TIME Magazine Corpus, and the Corpus of Historical American English – were used extensively in this thesis. The Corpus of Contemporary American English and the British National Corpus were also consulted through the BYU interface for comparative purposes. Some of the BYU corpora are monitor-like corpora, such as the Corpus of Contemporary American English (on which see Davies 2010), the News on the Web Corpus, and the Global Web-Based English Corpus. But some of the BYU corpora exhibit characteristics of both sample and opportunistic corpora.

The Soap Opera and TIME Magazine corpora used in Articles 1 and 2, for example, are clearly somewhat opportunistic; however, they were also compiled with linguistic research in mind, and are part-of-speech tagged, facilitating searches for syntactic constructions such as the passive. Their division into different time periods also makes them sample-like, as it is to some ex-

---

[29] An example of a study in which a monitor corpus is used is Deignan & Potter (2004), in which the English-language material in a typological study of body metaphor and metonymy is from the BoE.

tent possible to make diachronic comparisons of features in the material. COHA has an added sample-like dimension in that, in addition to a diachronic division of the material, it also includes material from four different genres. This facilitates a degree of cross-genre comparison and increases the corpus's (admittedly imperfect: see section 3.1) representativeness.

Table 1 shows the genres from which the passive tokens for the four articles were drawn. Three corpora and five genres are represented. I am not counting the TIME Magazine and COHA magazine data as separate genres, especially since TIME is one of the publications included in the popular magazines section of COHA.

Table 1. *BYU genres that were drawn on for primary material in the thesis studies.*

Corpus	Genre	Article that material was primarily used in	Time span investigated in study	Word count for study[30]
The Corpus of American Soap Operas	dramatic dialogue	Article 1	2002–2010	15,010,327
TIME Magazine Corpus	one popular magazine	Article 2	1923–2006	55,879,006
COHA	fiction	Articles 3 & 4	1870–1999	47,480,181
COHA	popular magazines	Articles 3 & 4	1870–1999	23,572,205
COHA	newspapers	Articles 3 & 4	1870–1999	10,171,402
COHA	non-fiction books	Articles 3 & 4	1870–1999	12,518,009

The five genres given in Table 1 are not representative of all American English, but they are different from one another in a few noteworthy respects. One of these is the orality-literacy cline investigated at length by Biber (1995) and Biber & Finegan (1989, 1992). Of the genres included in the studies, the soap opera transcripts used in Article 1 stand out as consisting solely of dramatic dialogue, which Biber & Finegan (1997) have found to exhibit characteristically oral linguistic features. However, there are important differences between the other genres in the studies as well. Fiction and academic prose of the type found in COHA's non-fiction books behave very differently from one another with respect to informational vs. involved production, elaborated vs. situation-dependent reference, and especially abstract vs. non-abstract style according to Biber & Finegan (1989). Finally, although newspapers and popular magazines are highly informational genres, which makes them similar to the academic prose of non-fiction books, they are also subject to the market forces which encourage colloquialization

---

[30] In Article 4, prepositional passives were searched for in a narrower diachronic range of COHA material. The word counts of the sections used in these searches are as follows: Fiction 7,936,303; Popular magazines 3,940,750; Newspapers 1,133,476; Non-fiction books 3,230,578. See Article 4, section 2 for details.

(Hundt & Mair 1999), which means they are likely to incorporate spoken-language features earlier than non-fiction books are. Hundt & Mair (1999) call such written genres "agile," whereas those which resist colloquialization longer (such as non-fiction books) are termed "uptight."

While not representative of the entirety of American English, these genre differences in orality/literateness and agility/uptightness offer good opportunities to trace the pathways of diachronic change in the use of passives. In Article 3, especially, it is apparent that the frequency of GET-passives increases in fiction or the agile genres of newspapers and popular magazines before they become acceptable in non-fiction books. Conversely, BE-passives remain frequent in non-fiction books after they have already begun to decrease in the more involved, situation-dependent, non-abstract fiction genre and then the more agile written genres. In addition, the strikingly high frequency of GET-passives in the material from the more oral Soap Opera Corpus compared to the more literate corpora gives an indication of just how frequent this construction might be in spoken American English.

The BYU corpora are not sampled in the same principled manner as many other digital language corpora. At this point, it becomes necessary to address an ongoing debate between corpus linguists over the relative usefulness of the large BYU corpora and the smaller, more carefully compiled Brown family corpora. The most recent manifestation of this "debate" is found in two adjacent chapters in *The Oxford Handbook on the History of English*: Davies (2016) and Hundt & Leech (2016). The authors respectfully take sides in a debate over the usefulness of large versus small corpora, and end up productively highlighting for the reader how each kind of corpus might best be used. It is my opinion that *all* of these corpora are worth exploring for what they have to offer, and that the research question should dictate the choice of corpus.

Had I only been interested in the decline of the BE-passive (see sections 2.1.1 and 6.1), I could have used Brown and Frown. Their small size would have been unproblematic had I been interested in tracking a construction that is so frequent in written English. The fact that new synchronic snapshots are being added to the Brown family (i.e. there will now be five points in time rather than two) will increase the validity of diachronic findings using these corpora. Hundt & Leech (2016: 179–180) point out that the tagging of the Brown corpora is more reliable than the tagging of the BYU corpora, because their small size allows for manual post-editing that ensures a high level of confidence in their correctness.

However, the BYU corpora, especially COHA, have the advantage of size when it comes to infrequent constructions such as the GET-passive and the prepositional passive. Hundt & Leech (2016: 178) recognize this:

> These [COCA and COHA] greatly exceed the BROWN family not only in size but also in number of data points (every decade or every five years rather

35

than every 30 years) and (in the case of COHA) in historical span–nearly 200 years rather than the century represented by our corpora […]. It is true that the BROWN family corpora are too small for studying change of low- and medium-frequency items, so their major use is in comparisons of high-frequency grammatical items and in the added level of regional variation, thus complementing corpora such as ARCHER for long-term diachronic change in the two major reference varieties of AmE and BrE.

Davies (2016: 172) also emphasizes the usefulness of his corpora in studying low-frequency constructions, writing:

The Corpus of Historical American English proves that it is possible to create corpora that are 100–400 times as large as these smaller corpora, but it is still possible to have corpora that are textually accurate, well-annotated, and genre-balanced.

While the point about the size of the BYU corpora being useful in studies of low-frequency items over a long diachronic span does not seem to be up for debate, Davies' last points about textual accuracy, annotation, and genre balance are. Mair (2006b) and Hundt & Leech (2016) are advocates of the "small is beautiful" corpus approach; Mair (2006b) specifically questions the role large corpora will be able to play in investigating GET-passives. These small-corpus advocates rightly point out that a major drawback of large corpora is that when there is so much material that it cannot be gone through manually, the precision of the tagging must be called into question.

It is certainly true that COHA is too large for manual POS-tagging, and even manual post-editing of computer tagging. It is also true that the tagger returns erroneous results. In my studies, the random data samples that were collected for further analysis were always gone through manually to ensure (a) that no false positives were included in the final counts and (b) that normalized frequencies of constructions were extrapolated from counts of correct examples. However, a number of valid tokens are likely to have been missed in the data collection due to corpus tagging error.

The genre balance of the Brown corpora must also be considered superior to the genre balance of COHA (Hundt & Leech 2016: 178–179). The Brown corpora are more rigorously sampled, include a wider array of genres, and are much more comparable with each other than the years of COHA are. The snapshots of the Brown corpora are sampled along exactly the same lines: the years of COHA are not. For instance, there is no newspaper material in COHA for the first 50 years of the time span covered by the corpus. Genre word counts for the few genres that are represented in COHA are not constant throughout the corpus, and some publications are unevenly represented. In my Articles 3 and 4, this genre imbalance was mitigated by considering extrapolated normalized frequencies per genre rather than percentages. Furthermore, all information on publications in COHA is available such that it is

possible to account for outliers and interpret results per genre in a principled, replicable manner.

The Soap Opera and TIME Magazine corpora, used in Articles 1 and 2 respectively, are more problematic in that they only include language from one genre (soap opera transcripts) and one publication (TIME Magazine), respectively. In both of these articles, results were compared with results from other corpora, and the findings were supported on both occasions. Thus, there is no real reason to avoid using these large, one-genre corpora, as long as results are checked against other sources and conclusions are responsibly drawn.

Lastly, a final word on the quest for representativeness: as I said in section 3.1, genre balance is not equivalent to representativeness of a speaker population (nor are Hundt & Leech 2016 trying to make any such claim). The genres included in the Brown family are meticulously comparable *with one another*, but they are not representative of all American or British English, even so-called Standard American or British English, for the reasons given in section 3.1. While the wider range of genres in these corpora might make them *more* representative than COHA, it would be inaccurate to confidently claim that these beautifully compiled, smaller corpora are entirely representative of a speaker population, despite their greater comparability and more uniform balance of different publications.

For the purposes of my study, then, the increased diachronic span and the large size were of paramount importance in corpus selection; this is why the BYU corpora were used. For other research questions, smaller, more rigorously sampled and tagged corpora might be a better fit. As electronic resources become more sophisticated, it is likely that the range of linguistic corpora will grow, and it will be up to linguists to make responsible, replicable use of them.

3.3. Searching for passives

For all four of the studies included in this thesis, I searched for passive sentences using the web interface that is freely available for the BYU corpora. The corpora that were used in this thesis are all tagged for part-of-speech, which meant that I could search for the auxiliary lemma (either BE or GET) followed by a past participle. The basic searches performed for all studies were:

auxiliary lemma + (optional adverb) + past participle[31]

To give some idea of the size of the BYU corpora, I offer the example that a search for BE followed directly by a past participle in all of COHA returns

---

[31] The method for retrieving prepositional passives entailed the same basic search for passives with some additions. See the methods section of Article 4 for details.

3,250,489 hits. For the 1910s, one of the decades under consideration in Articles 3 and 4, there are 188,622. Obviously, it would not have been feasible to go through every single instance of auxiliary + past participle manually as it would have been with a smaller corpus (see section 3.2).[32] However, it was absolutely necessary to find some method of manually examining search returns, since searches inevitably result in a number of "false hits." In fact, only about half of the search returns overall were classified as true, central passives according to the criteria I lay out in section 4.1.

Examples of sentences returned by my searches are given in (22)–(27). Of these example sentences, only (22) and (23) qualify as central passives; the searches obviously also return sentences like (24)–(27), which are either ambiguously passive or not passive at all (see section 4.1 for a discussion).

(22) For a time, evidently, the grand life <u>got lived</u> here. (COHA Fiction 1993)

(23) All these plants have a genuine medicinal value, and many of them <u>are still used</u> as cures in country districts to-day. (COHA Non-fiction 1953)

(24) Moreover, the attempt at social criticism <u>is strained</u>. (TIME 1960)

(25) Becky started to play two years ago; both are aware that women can never be as good at the game as men but that doesn't prevent them from <u>getting enormously excited</u> about it. (TIME 1928)

(26) He's not gonna rest until he <u>gets rid</u> of David. (SOAP AMC 2010)

(27) She's also the baby of the family, so she<u>'s used </u>to getting her way. (SOAP OLTL 2002)

The search interface of the BYU corpora also offers a function which generates randomly selected subsets of the search returns. This allowed me to harvest large, random sets of tokens for further study. The proportions of central passives in these randomly selected sets allowed me to extrapolate, in a principled manner, normalized frequencies of passives in the corpora. The method was as follows:

1. Subsets of several hundred randomly selected tokens were collected from the desired corpus, genre, and time period.[33]
2. These tokens were analyzed for centrality (section 4.1).
3. The percentage of central passive tokens was calculated.

---

[32] For GET + past participle, which is much rarer, it actually sometimes *was* possible to go through every instance in a particular year, decade, or genre. For BE + past participle, it never was. See the individual articles for details.

[33] These obviously varied from one study to the next. Again, for GET + past participle and for the prepositional passive, which are much rarer than BE + past participle, I sometimes collected *all* instances for further study. See the individual articles for exact figures.

38

4. This percentage was multiplied by the number of total tokens returned by the original search. This yielded the estimated raw frequency of central passives for the relevant decade/year/genre.
5. The normalized frequency of central passives was extrapolated using the corpus word count for the relevant decade/year/genre and the estimated raw frequency that had been calculated in step (4).

This method offered a principled, replicable means of studying developments in the frequency of use of the passive constructions, a question that was under investigation in all of the four articles. The random subsets of central passives that were collected using this methodology also constituted the primary data in which I examined developments along such parameters as adversativity, situation type, and thematic role, which in turn led to insights into the characteristics and development of the different passives.

4. Classification schemes

One possible contribution of this thesis to future research is the replicable syntactic and semantic classification methodologies on centrality (developed for Article 1, used in all articles), adversativity (developed for and used in Article 1), situation type (developed for Article 2, also implemented in Article 3), and thematic role (Article 4).[34] As indicated, the classification system pioneered in one article paved the way for work in subsequent articles. For example, it was necessary to have established a reliable method of detecting central passives (section 4.1) before further work on situation type and thematic role (sections 4.3 and 4.4) could be undertaken in Articles 2, 3 and 4. The adversativity classification described in section 4.2 was only used in Article 1, but has already been fruitfully implemented by another researcher (Coto Villalibre 2015: 18).

4.1. Centrality

Not all examples of passive auxiliary[35] + past participle meet the criteria laid out by Quirk et al. (1985: 167–171) to qualify as true, "central" passives. The "passive gradient" which Quirk et al. describe in helpful detail (1985: 167–171) formed the basis of my "centrality tests," as described at greatest length in Article 1 (Schwarz 2015: 158–159).

The test questions, reproduced in (a)–(f) below, are used to separate instances of BE/GET + past participle into the categories 'central passive,' (where only a verbal, passive reading of the sentence is possible) 'semi-

---

[34] Earlier versions of the centrality and adversativity classification methodologies were developed for Nilsson [Schwarz] 2012 and refined for the PhD thesis.
[35] See section 2.1.2 on the status of GET as a passive auxiliary.

passive,' (where the participle is ambiguously verbal or adjectival), and 'pseudopassive,' (where the sentence bears only a surface resemblance to a passive).[36]

(a) Is there an active analogue?
(b) Can the participle be co-ordinated with an adjective?
(c) Can the participle be modified by *quite* or *rather*?
(d) Can BE or GET be replaced by a lexical copular verb like FEEL or SEEM?
(e) Is there no possibility of agent addition?
(f) Is the active version of the sentence perfective?

Central passives, which were the constructions in focus in this thesis, are those for which the answer to (a) is "yes," but the answer to *all other questions* is "no," as in (28) and (29).

(28) When she collapses, **Miss Crawford** <u>is taken</u> to a psychopathic ward. (TIME 1940s)
(29) Wonderful things happen (sometimes) when **a painting** <u>gets stolen</u>. (TIME 1960s)

If the answer to question (a) was yes, but the answer to at least one of (b)–(d) was *also* yes, then the sentence was found to be a semi-passive, as in (30) and (31).

(30) **Some of the passengers** <u>got so frightened</u> that an itinerant preacher on the foredeck judged this to be an auspicious moment to hold a revival meeting and to circulate a plate. (COHA Fiction 1951)
(31) It dawned on Julian that MacGregor was really concerned about him. **He** <u>was touched</u>. (COHA Fiction 1993)

Finally, pseudopassives, those sentences which bear only a surface resemblance to a passive sentence, were identified if the answer to (a) was "no" and at least one of (e) or (f) was "yes." Obviously a negative answer to (a) was theoretically enough in and of itself to classify a sentence as pseudopassive, but having a battery of test questions was often helpful in making difficult calls on naturally-occurring language data. Examples of pseudopassives are found in (32) and (33).

---

[36] The term pseudopassive is a potentially confusing one, as some linguists use it to mean "prepositional passive." In this respect I follow Quirk et al. (1985: 169–170) in using pseudopassive to mean "a construction that looks like a passive but isn't."

(32) We have to pull together now. Otherwise, **this family** is done. (SOAP AMC[37] 2010)
(33) **You**'re just gonna have to get used to that. (SOAP BB[38] 2010)

After writing the first article, where I considered GET *married* to be a central passive (but presented frequencies both with and without that particular passive), I decided that GET *married* needed to be an exception to the centrality tests. While GET *married* technically passes the test for centrality, it is so frequent, and seems so often to be adjectival, that I ultimately decided to follow Huddleston & Pullum (2002: 1441) and Leech et al. (2009: 155–156) in treating it as a semipassive unless a *by*-agent is present, as in (34) and (35).

(34) Don't ask David again. We don't need his consent. **We**'ll be married *by the elders* in Bethlehem. (COHA Fiction 1959)
(35) **She and George** got married in January *by the justice of the peace* because he couldn't wait, and she could think of no reason why he should. (COHA Fiction 1999)

In a way, it would seem to make intuitive sense that a study on passive voice should focus on the sentences that qualify as true, central passives; this is mostly what has been done in the four articles in this thesis. However, the inability of writers to reliably identify passive sentences (Pullum 2014: 73) means that if change in the passive is happening above the level of consciousness, constructions that may be *seen* by writers as passive, especially semipassives, should possibly be taken into account in studies of prescriptivism and colloquialization. Furthermore, semipassives might share in the grammaticalization trajectory along which central GET-passives are proceeding. The question of whether or not to examine semipassives was touched on in Article 3, where the frequency trajectories of semipassives and central passives with GET were compared. The implications of those trajectories for the further grammaticalization of the GET-passive are discussed in Article 3 and in section 6.2 of this introductory survey.

4.2. Adversativity

In section 2.1.2, I discussed the supposed adversativity constraint on GET-passives, as described in Chappell 1980; Quirk et al. 1985: 161; Biber et al. 1999: 481; Carter & McCarthy 1999; and Huddleston & Pullum 2002: 1442. Leech et al. (2009: 157) find that, while not universally adversative, GET-passives tend to have non-neutral semantics in the corpus data used in their study (the 1960s and 1990s sections of the Brown family of corpora: the

---

[37] AMC = All My Children
[38] BB = The Bold and the Beautiful

only decades available at that time). The adversativity of the GET-passive as compared to a comparable set of BE-passives was one of the research topics explored in my Article 1. In it, I employed a test for adversativity based on Persson (1990: 52), which entails asking the question "Is it generally thought worse to be X than not to be X?" By way of example, let us consider sentences (36)–(38).

> (36) Damn. If I talked to my parents like that, **I**'d get slapped. (SOAP OLTL[39] 2010)
> (37) **She** got paged to an emergency in the E.R. (SOAP ATWT[40] 2010)
> (38) Well, **I** get paid handsomely to do that. (SOAP YR[41] 2010)

For sentence (36), it is generally thought worse to get slapped than not to get slapped, so the sentence is coded as adversative. Sentence (37) does not have a clear adversative or benefactive sense; nothing in the situation in this line or the surrounding lines indicates that it was particularly worse to be paged to the E.R. than not, or better to be paged to the E.R. or not. Sentence (37) is thus coded as neutral. Finally, since it is generally thought better to get paid than not to get paid, sentence (38) is coded as benefactive.

The adversativity test is somewhat subjective, as are the tests for situation type and thematic role which are described in 4.3 and 4.4. The best a researcher can hope for is (a) that their examples indicate how they have reasoned in their analysis, and (b) that they have aimed for internal consistency, as I have tried to do here.

4.3. Situation type

Not satisfied with the ability of tense, aspect, and mood to adequately describe the "temporal semantics" that a verb confers on a sentence, Vendler (1957) introduced the concept of situation type, also known as *Aktionsart*. The system has been much refined since Vendler's introduction; the classifications I used in this thesis were loosely based on Quirk et al. (1985: 200–209). I did not make use of the entire gamut of situation types Quirk et al. suggest – only those which seemed to pertain to my passive sentences. First, the passives were divided into dynamic and stative sentences. Stative passives are very rare, and the test for them is simple. If the passive sentence *very strongly resists* having its active version put into the progressive, the sentence is considered stative. Example (39) shows such a sentence.

---

[39] OLTL = One Life to Live
[40] ATWT = As the World Turns
[41] YR = The Young and the Restless

> (39) The tragedy made a tremendous difference to Niki's father, **who**
> **was universally known** as Papa Balinski. (COHA Non-fiction
> 1990s)

A progressive version of the relative clause in (39) might read *who people*
were universally knowing as Papa Balinski. This is highly unacceptable,
indicating that this is a rather clear example of a stative passive. Stative pas-
sives were a marginal category, and were not classified further along Quirk
et al.'s (1985: 201) subcategories; however, almost all of them would have
been examples of "states."

Most of the passive sentences in my studies were dynamic, which is to be
expected. As opposed to stative passives, dynamic passives are identified by
the fact that their active versions *can* be put in the progressive; however,
whether their active, progressive versions have the same semantics as the
non-progressive version determines what kind of dynamic sentence they are.
Following Quirk et al. (1985: 201) and Smitterberg (2005: 162), dynamic
sentences were further classified as either durative or punctual, and either
conclusive or nonconclusive (often called telic/atelic). Vendler (via Brinton
1988) and Quirk et al. (1985:201) also suggest volition/agentivity as a pa-
rameter in the classification of situation type. However, Brinton (1988: 32)
argues that agentivity should be a distinct category. Following Smitterberg
(2005: 162), I not only consider agentivity separately, but actually find it so
subjective to determine that I content myself with looking at human vs. non-
human NP complements as the separate parameter.[42]

The parameters under consideration for dynamic passives in this thesis, as
adapted from Quirk et al., are shown in Table 2. The names of the situation
types are taken from Quirk et al.'s (1985: 201) "agentive" subcategories, as
these correspond most closely to the original situation types proposed by
Vendler (1957) that are most frequently referred to in the literature.

Table 2. *Situation type categories for dynamic passives used in the thesis, as*
adapted from Quirk et al. (1985: 201).

Accomplishment	Activity	Transitional Act	Momentary Act
[+durative]	[+durative]	[-durative]	[-durative]
[+conclusive]	[-conclusive]	[+conclusive]	[-conclusive]
The house finally got built.	*He's being taken care of by his niece.*	*He got fired.*	*It frequently got mentioned by stu-dents.*

---

[42] The classification for human vs. non-human NPs was straightforward enough that it does
not merit its own section in this introductory survey. For Article 4, I followed Bosse,
Bruening & Yamada (2012) in classifying NPs as either sentient vs. non-sentient. See their
paper and Article 4 for a discussion of the distinction.

The durativeness and conclusiveness of each sentence was also tested by putting the active version of the passive sentence into the progressive. When the meaning of the sentence was essentially unchanged (aside from the change in aspect), the sentence was considered durative, as in (40) and (41). Where the progressive imbues different temporal semantics on the sentence (that the action immediately precedes a change of state, as in (42), or that the action becomes iterative, as in (43)), the sentence is punctual.

The conclusiveness of the sentence is determined by asking the question proposed by Garey (1957: 105): "If one was *verb*ing, but was interrupted while *verb*ing, has one *verb*ed?" For nonconclusive sentences such as (41) and (43), the answer is "yes." For conclusive sentences, as in (40) and (42), the answer is "no." Articles 2 and 3, in which situation type is in focus as a research question, detail these tests even more extensively and offer further examples.

(40) Accomplishment [+durative], [+conclusive]

The article in question <u>may have been written</u> by Dr. Milman himself, who was then one of the principal contributors to the great Tory periodical, and he perhaps had read the poems, but apparently without much edification. (COHA Magazines 1876)

(41) Activity [+durative], [-conclusive]

No matter what my husband and I do to protect and prepare Chelsea, **her future** <u>will be affected</u> by how other children are being raised. (COHA Non-fiction 1996)

(42) Transitional Act [-durative], [+conclusive]

Highly paid players <u>get traded</u>, and eventually everybody retires because they're through when they can't produce any longer. (COHA Newspapers 1957)

(43) Momentary Act [-durative], [-conclusive]

Her critics say she has been little more than a caretaker, feeding off the momentum created by her predecessor, **Tom McEnery, who** <u>is widely credited</u> with transforming San Jose into a modern big city and with reviving the downtown neighborhood. (COHA Newspapers 1997)

The classification for situation type was one of the most difficult endeavors of this thesis. Smitterberg (2005: 165), who used a slightly different set of situation-type categories in his corpus study of 19th-century progressives,

describes the difficulty of imposing theoretical categories on naturally-occurring language data:

> To some extent, Aktionsart categories constitute an idealized representation of reality, which sometimes leads to difficulties regarding the classification of situations. For instance, Comrie (1976: 41 ff.) points out that it is difficult to find situations that are strictly [-durative], or punctual, in that they completely lack duration, although this does not invalidate the recognition of punctuality as an Aktionsart category.

Comrie (1976) specifically identifies the punctual categories as problematic, and I emphatically agree. Not only is it difficult to decide just how much duration counts as duration (although the test questions described above do help), Quirk et al.'s four-way distinction (as was shown in Table 2), while appealingly symmetrical, does not, in retrospect, seem as useful as the original categories proposed by Vendler, which are less detailed when it comes to the punctual categories of dynamic sentences. Vendler's typology, modified from Smitterberg (2005: 160), is shown in Table 3.

Table 3. *Vendler's Aktionsart typology for dynamic sentences (via Brinton 1988: 28f. and Smitterberg 2005: 160)[43]*

Activity	Accomplishment	Achievement
[+durative]	[+durative]	[-durative]
[-telic]	[+telic]	-

The first two of Vendler's durative categories correspond to Quirk et al.'s (1985: 201): Activity [+durative], [-conclusive] and Accomplishment [+durative], [+conclusive]. However, there are no separate telicity (conclusiveness) categories for punctual sentences; instead, the categories of Transitional and Momentary Acts are both subsumed under Vendler's Achievements, which are simply [-durative]. In retrospect, I think Vendler's original categorization makes better intuitive sense, and I might have preferred to have removed the category of Momentary Acts from my classification system. Sentence (43) was intentionally selected as a particularly troublesome example. The active, progressive version of (43) would be something like *who people are widely crediting with transforming San Jose into a modern big city*. It is the iterative, over-and-over again sense of *crediting*, imbued by the plural nature of the implied agents of the action and the adverb *widely*, that leads to this sentence's categorization as a Momentary Act according to the criteria. However, such sentences may seem more intuitively to belong to the Activity category. This classification system may require further refinement before being implemented in future studies.

---

[43] Brinton (1988) and Smitterberg (2005) use the term *telic* where Quirk et al. (1985) use *conclusive*.

Situation type classification, in sum, was onerous and somewhat subjective. However, by rigorously adhering to replicable tests with as high a degree of internal consistency as I could muster, I hope to at least have achieved results that are diachronically comparable across genres for the purposes of the studies in this thesis. In my future work, in which I hope to analyze the telicity of prepositional passives (see Article 4), I may remove the "iterative" test which leads to the placement of sentences such as (43) into the Momentary Act category.

4.4. Thematic role

Article 4 focused on the prepositional passive; specifically, on the characteristics of this typologically rare construction which allow for the promotion of a prepositional object to subject position. As discussed in section 2.1.3, the agreement in the literature is that there is an "affectedness" condition which permits passivization: Bolinger (1977: 67) wrote that a passive, whether prepositional or not, is only possible when "the subject […] is conceived to be a true patient, i.e., to be genuinely affected by the action of the verb." The phrase "true patient" suggests that thematic role might be a good means of exploring the data, especially as traditional affectedness "tests" of the kind summarized in Beavers (2011) were not really applicable to naturally-occurring language data, having been invented for explanatory models (on which see Article 4).

In Article 4, affectedness testing was carried out by first coding the prepositional passive sentences for thematic role of subject, and then proposing that certain thematic roles are more affected than others. Just as with adversativity and situation type, the coding for semantic role was both difficult and (unavoidably) somewhat subjective. In fact, in his paper on thematic "proto-roles," Dowty (1991: 572) recognizes that semantic categories such as these are less clear-cut than syntactic and phonemic categories, and although I have again aimed for a high degree of internal consistency, I imagine that the classification of naturally-occurring language data into these roles is something that linguists could discuss at great length. This classification process was also rather more exploratory than for centrality, adversativity, and situation type, and the categories and groupings that are suggested for this study could certainly be problematized and refined for future work.

In Article 4, I propose that the following five thematic roles in the data are the most affected by the verb:

- Patients: which undergo real change, as in (44)
- Themes, which move or are located somewhere, as in (45)
- Results, which come into or go out of existence, as in (46)
- Experiencers, as in (47)
- Recipients/Beneficiaries, as in (48)

46

(44) **This boat, the Wilson Small**, is disabled. **She** <u>was twice run into</u> today […] (COHA Non-fiction 1959)

(45) **The totality of the labourer's self**, and not just a detachable object in his possession, <u>is parted with</u> and transferred to somebody else's control. (COHA Non-fiction 1990)

(46) Then **this whole cursed business** <u>would be done away with</u>. (COHA Fiction 1910)

(47) "But," Ford insisted in hurt tones, as though **he** <u>were being trifled with</u>, "you have been told you look like him, haven't you?" (COHA Fiction 1910)

(48) **She** <u>was written to</u> by Dr. Coffee, Eye Specialist. (COHA Non-fiction 1910)

Interestingly, these five affected roles only accounted for about half of the prepositional-passive data and about two-thirds of the non-prepositional-passive data. Most of the remaining sentences conferred what I took to be the role of "percept" on the subject-referent (Saeed 2003: 151–152). Saeed defines percept as "the entity which is perceived or experienced." As such, it bears some similarity to the more familiar role of Stimulus. The difference is that the inclusion of the word "perception" allows for intentional involvement on the part of the perceiving participant, as in sentences (49) and (50).

(49) In England **every great question** <u>is looked upon</u> as fundamentally settled. (COHA Magazines 1870)

(50) John felt a thud of pain centering in a core of numbness where his right elbow used to be, and what happened, he realized, was that **his head** <u>had been aimed at</u> and that the bullet, going slightly wide of the mark, had found a target in his elbow… (COHA Fiction 1959)

Perceptually salient subjects were taken as an indication that Davison (1980: 54) and Huddleston & Pullum (2002: 1446) were correct in suggesting that prepositional passives are also possible where either "a significant property" of the subject-referent is expressed, or where the subject referent has perceptual salience. The thematic-role classification thus revealed interesting properties of prepositional passives as regards not only affectedness, but perceptual salience of subject-referent as well.

Lastly, the prepositional-passive data included the very marginal category of instrument, most of which were abstract rather than concrete, as in (51).

(51) **Painting** <u>is seldom resorted to</u> for merely decorative effects, but rather when a high degree of visibility is mandatory. (COHA Non-fiction 1950)

47

The data actually contained no truly locative subject-referents, indicating that this is a vanishingly rare thematic role in prepositional passives.

5. Individual article summaries

This section includes summaries of the individual articles which, together, form the doctoral thesis. Separately, the articles, which were written in order such that the findings of one could inspire the focus of the next, proceed in a natural way along at least four dimensions: time (they move into the past diachronically from the first article onward), classification schemes (those developed for one article can be implemented as background for the next), text type (the scope becomes broader and goes from informal to formal), and focus (from a broad focus on both GET- and BE-passives, to GET-passives only, to prepositional passives).

5.1. Article 1: Passive voice in American soap opera dialogue

In Article 1, both BE- and GET-passives are investigated in very recent, speech-related material. Under specific investigation are their frequencies over the first decade of the 21st century, which are compared with findings on earlier time periods in other studies, and whether or not they are used in an adversative sense, as this has previously been claimed for the GET-passive but not the BE-passive. In this study, I attempt to get at recent, speech-like data and make use of the Corpus of American Soap Operas (Davies 2012), which includes 100 million words of transcribed soap-opera dialogue from the years 2001–2012. The benefits and drawbacks of this unusual material, one of which has to do with the remarkably high incidence of *get married*, are discussed at length in the article.

Of the articles in this thesis, it is this article that contains the most detailed account of how passives can be retrieved from a tagged corpus. A replicable methodology for identifying what Quirk et al. (1985: 167–171) call "central" passives as opposed to "semipassives" and "pseudo-passives" is presented here. The method is applied to the data sets. The data are also tested for adversativity using Persson's (1990: 52) test. It is found that (a) there is a decrease in the frequency of BE-passives, but no change in the frequency of GET-passives and (b) BE-passives show a surprising amount of diachronic fluctuation as regards adversativity. In the most recent data, GET-passives are no more adversative than BE-passives.

The GET-passive's frequency in the Soap Opera Corpus, while stable, is much higher than in less speech-related corpora, and it is suggested that the GET-passive may have reached "saturation point" (Leech et al. 2009: 210) in spoken American English. The GET-passives in the data are likely to be ad-

48

versative, which does not indicate semantic bleaching; but the fact that the BE-passives are as adversative as GET-passives in the most recent data shows how important it is not to draw conclusions about GET-passives without comparing them to BE-passives. It is also suggested that adversativity may be a function of genre; this, however, does not account for the fluctuation observed in the BE-passive.

The declining frequency of BE-passives even in recent, speech-related data called for further investigation. Since prescriptive influence is a likely reason for their decrease, even in soap-opera transcripts, the frequencies of two other proscribed linguistic features, namely *which* in restrictive relative clauses and stranded prepositions, are also examined. There is no diachronic change in their frequencies. However, while a high instance of preposition stranding suggests that writers are not responding to that particular proscription, the very high proportions of restrictive relative *that* over *which* indicate that prescriptive influence cannot be discounted as a factor in the material. The marked decrease in the frequency of BE-passives over the short time period studied also indicates prescriptive influence and/or colloquialization.

5.2. Article 2: "Like getting nibbled to death by a duck": Grammaticalization of the GET-passive in the TIME Magazine Corpus

In this article I make use of diachronic corpus data to study some key syntactic and semantic characteristics of GET-passives. The hypothesis for the paper is that changes in the tense, aspect, and/or situation type (*Aktionsart*) of the GET-passive in the data might reveal that it has been undergoing morphosyntactic generalization, one of Hopper & Traugott's (2003: 104–106) stages in grammaticalization. The corpus material included in the TIME Magazine Corpus differs from the corpus material for Article 1 in two important ways: the TIME material is (1) slightly earlier, covering most of the 20th century, and (2) an "agile" (Hundt & Mair 1999) written genre which can be assumed to be moving in the direction of spoken-language due to market forces. If the Soap Opera Corpus data suggest "saturation point" in the frequency and use of the GET-passive in spoken language, data extracted from the TIME corpus might show stages of grammaticalization that had led up to that point (which would be delayed in the written language).

The data are analyzed along three parameters that have not, to my knowledge, been systematically studied for GET-passives in previous works. The results for the study of tense and aspect of GET- vs. BE-passives show some differences between the two constructions, but not much in the way of diachronic development (and thus grammaticalization) of GET-passives. Situation type, on the other hand, turns out to be the most interesting variable under consideration in Article 2: the GET-passive's preference for Transi-

tional Acts (which makes intuitive sense in light of Fleisher 2006) seems to weaken over time, indicating morphosyntactic generalization along this parameter such that, in the most recent data, the use of the GET-passive more closely resembles the fully-grammaticalized BE-passive in terms of temporal semantics. A mini-study of the Fiction and Newspaper genres of the Corpus of Historical American English (COHA) (Davies 2010) supports this result.

5.3. Article 3: Signs of grammaticalization: Tracking the GET-passive through COHA

The situation-type findings of Article 2 are followed up across a longer diachronic span and across genres in Article 3. Again, GET-passives are examined against a control group of BE-passives. The expected frequency trajectories (decline in BE-passives, increase in GET-passives) are evident, and the frequency changes across individual genres suggests colloquialization, as both the decline in BE-passives and the increase in GET-passives start earlier in the less formal genres before spreading to the most formal ones. Furthermore, the "central" passive is shown to be the construction undergoing the most change in the data, justifying a focus on the true, verbal passive of linguistic description rather than all constructions that writers might (even mistakenly) identify as passive, and suggesting that grammaticalization of the central passive function is the reason for the increased frequency of the GET-passive.

Collected subsets of GET- and BE-passives are then analyzed along the parameters situation type, subject type, and past participle lexeme. The situation type findings for COHA are reminiscent of those from the TIME corpus for Article 2, but it is noted that the important shift in the use of the GET-passive to new situation types which is visible in both studies is not necessarily in Transitional Acts vs. all other categories, but in conclusive (telic) vs. non-conclusive (atelic) situation types. The subject-type investigation also indicates a weakening of the restriction of GET-passives to human subjects, indicating that the agentivity requirement proposed in the literature may have been weakening over the time span in the study. Finally, the range of past participles which are used with passive-auxiliary GET broadens over the time period, indicating that the GET-passive has become increasingly productive and is not restricted to as limited a set of lexicalized types now as it once was. Taken together, the results of the study show clear indications of further grammaticalization of the GET-passive in the COHA data.

5.4. Article 4: *This must be looked into*: A corpus study of the prepositional passive

The final article in the thesis focuses on the prepositional passive, which, while a favorite topic for generativist studies, has been the object of relative-

ly little empirical study. Only BE-passives are included here, as the prepositional passive with GET is found to be too rare in the collected data to allow for generalizations to be made.

In Article 4, the material in COHA is used to investigate theoretical claims about the prepositional passive against a control group of non-prepositional passives. This control group consists of the BE-passives that were collected for Article 3, with the prepositional passives (only 1% of the total) removed. The material for the article thus consists of two sets of BE-passives (one prepositional and one "regular"), spanning the years 1870–1999; however, in this study, the only observable change is in frequency of use: the prepositional passives decline at the same rate as non-prepositional BE-passives do. The discourse and semantic preferences of the prepositional passives are stable over the entire time period included in the study.

The parameters under examination in the data, all of which are suggested by the existing canon of theoretical literature on prepositional passives, are the following: degree of lexicalization of the verb + preposition (operationalized by type/token ratio); participle + noun + preposition combinations with light verbs such as *taken care of* and *made use of*; coordination with another passive verb; affectedness; and perceived significance of subject-referent. A thematic role classification is implemented as a means of studying the last two of these parameters, and this classification highlights the difficulty of applying theoretical semantic categories to naturally-occurring language data. The empirical findings of the study largely support the prior intuitions expressed in earlier theoretical works, but the corpus results do occasionally call for nuancing of earlier claims, and suggest fruitful avenues for further research.

6. Findings

This section deals with overall findings of the thesis, i.e. an interpretation of the results of the four articles taken together. I focus here on relating findings to the aim of the thesis as stated in section 1: in section 6.1, I reflect on what the studies reveal about the how the passives differ from one another and how they develop diachronically in terms of their syntactic and semantic preferences. The more established, canonical BE-passive is largely treated as the unmarked construction against which GET-passives and prepositional passives are compared; however, the BE-passive is discussed in its own right at the end of section 6.1. In section 6.2, I explore how the evidence in the studies relates to theories that may play a role in changes in the use of the passives.

6.1. Differences between BE-, GET-, and prepositional passives

The BE-passive and the GET-passive are "competing" features in that they offer identical information-structural discourse functions. However, while the GET-passive seems to have been enjoying an immense surge in popularity over the time period studied, especially in the more recent decades, it cannot be considered a serious competitor to the BE-passive, and must still be regarded as a more marked and marginal member of the passive category. Even in the most recent, speech-related data from the Soap Opera Corpus, the GET-passive is only about a tenth as frequent as the BE-passive. The GET-passive is an informal feature associated with spoken language; as such, its increase in written genres is taken as a sign of colloquialization (sections 2.1.2, 2.2.2, 6.2). In this thesis, I also manually examined large sets of corpus data for signs that the GET-passive has been undergoing further grammaticalization as it has become more frequent in written language (in the full recognition that these written-language changes probably only reflect developments that took place much earlier in spontaneous spoken language).

Basically, grammaticalization of the GET-passive was investigated by testing whether any of the proposed "constraints" on the GET-passive (section 2.1.2) have been weakening diachronically. The constraints that were investigated in Articles 1–3 were: adversativity, agentivity (tested by looking for human subjects) and situation type. The adversativity findings were inconclusive; the GET-passive's association with non-neutral semantics does *not* seem to have weakened (see also Leech et al. 2009:157). However, in the most recent data from the Soap Opera Corpus (Article 1), the BE-passive showed similarly non-neutral semantics, and there was no statistically significant difference between the two kinds of passive (on which see the discussion of the BE-passive at the end of this section). This, however, may very well be due to the nature of the Soap Opera material, and bears looking into in other genres as well.

Along the agentivity and situation type parameters, the GET-passive showed signs of morpho-syntactic generalization. The GET-passive underwent a change away from use with human subjects in some genres of COHA (Article 3), which indicates possible weakening of the agentivity constraint; genre drift cannot be the explanation for the change in this case, as the proportion of BE-passives with human subjects was either stable or actually increased diachronically across COHA. It was also found in Articles 2 and 3 that the association of the GET-passive with telic situation types has been weakening over time. This is a strong indication that the GET-passive is grammaticalizing in such a way that it is becoming more similar to the canonical BE-passive in terms of temporal semantics.

Although not a constraint, the GET-passive also showed development in its increased use with a more varied set of participles. In the earlier data used

in Article 3, the GET-passive is mostly used with a restricted set of participles. In the most recent data for that article, there are instead a great many single-occurrence past participles, indicating that the GET-passive becomes more productive over the time span. Rodríguez-Puente uses the same reasoning on her diachronic study of phrasal verbs (2016: 90), finding that single-occurrence items increase diachronically and that "a measure of this kind is very appropriate with fairly large corpora," which is exactly what makes it so useful in Article 3.

For Articles 1–3, the data sets of BE- and GET-passives included prepositional passives, which are very rare (only about 1% of these data sets). However, for the purposes of Article 4, a large set of these rare passives was collected for further study in their own right. The prepositional passive with GET is so rare that it had to be excluded from this study, so only prepositional and non-prepositional passives with BE were examined. There are other differences between this study and the preceding three. For one thing, the only diachronic change in the prepositional passive was the expected decline in frequency found in all passives with BE, whereas the discourse and semantic properties of prepositional passives showed diachronic stability. Furthermore, the non-prepositional passives were *only* included as an imperfect control group (on which see Article 4 for details), with the true focus of the study being the empirical investigation of theorized properties of prepositional passives. The findings of Article 4 strongly support theoretical claims about the construction, and also suggest that these theories may hold for all prepositional passives and not just where the verb and preposition do not seem to constitute one, lexicalized constituent.

As regards the properties of the prepositional passive vs. the properties of the other kinds of passive, my claims must be tentative due to a lack of truly comparable data. However, it is noteworthy that both the prepositional passive and the GET-passive have been more likely to be used with human subjects than the BE-passive. It would be interesting to study the thematic roles of subject-referents in GET-passives, and the situation types of prepositional passives and see what else they might have in common. It may be the case that these rarer, more marked passive constructions resist neutral contexts more strongly and respond to similar constraints. Again, however, the prepositional passive has a different, and typologically rare, information structural function; it seems likely that prepositional passives will always have special properties, as it may "take more" for them to be able to constitute acceptable passives.

Finally, the control-group passive, the canonical passive, the oldest of the three: the BE-passive. The BE-passive constituted the baseline against which the other two constructions were measured, but it has also been undergoing major usage changes in its own right. Mostly, what is interesting about the use of the BE-passive over the time period in the studies is the dramatic de-

53

crease in frequency that will be discussed at length in terms of prescriptivism and colloquialization in section 6.2.

But there was change along some of the other parameters that were under consideration in the studies as well. For instance, as I stated above, the BE-passives in the Soap Opera Corpus underwent significant change in neutrality of semantics. While the GET-passive remained as non-neutral as ever in this very recent material, the BE-passive showed an as yet unexplained change from more to less neutral, such that it actually resembles the GET-passive in the 2010 data. Furthermore, the BE-passive showed an increased association with human subjects in the Newspaper genre of COHA in Article 3. Both of these findings were unexpected, and deserve some comment here.

I suggested at the beginning of this section that the nature of the Soap Opera material probably plays some role in the non-neutral semantics found in the passives in that genre. However, the dramatic nature of soap operas cannot account for the change that the BE-passives display over the brief time period. Neither can genre account for the increase in human subjects with BE-passives in the newspaper genre of COHA, since the GET-passives do not behave in the same way. It is possible that these findings are simply due to a random quirk in the corpora; but it is equally possible that the BE-passive is undergoing some changes as it becomes less popular. Perhaps, as the BE-passive falls out of fashion, its use will become increasingly constrained not only to certain genres but to certain semantic contexts. Time and future research will tell.

6.2. Language change and the passive in American English

Frequency findings in all studies were in line with earlier research (Leech 2004; Mair 2006a; Leech et al. 2009; Anderwald 2016) in that BE-passives
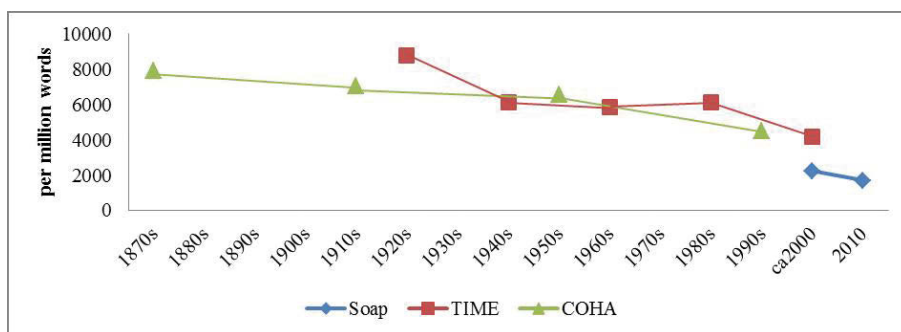

Figure 1. *Extrapolated normalized frequencies of BE-passives from the main corpora used in the studies.*
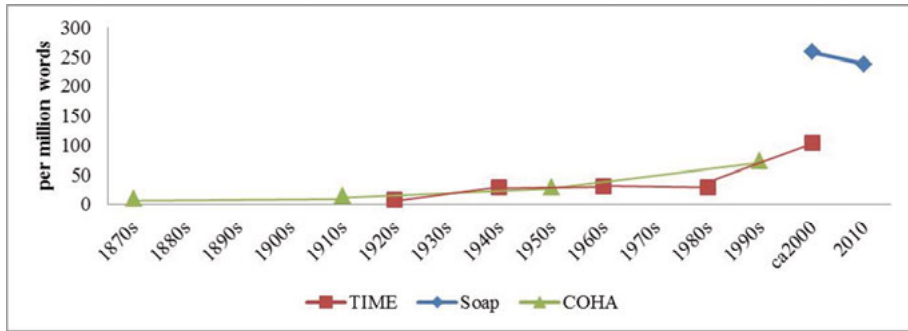
Figure 2. *Extrapolated normalized frequencies of central* GET-*passives from the main corpora used in the studies.*[44]

declined in frequency diachronically (Figure 1) while GET-passives increased (Figure 2). Also as expected, the GET-passive cannot be said to be in competition with the BE-passive (at least in the written genres in the studies), as the BE-passive is still ten times as frequent as the GET-passive even in the most recent, speech-related material available (the Soap Opera Corpus).

The genre perspective afforded by the BYU corpora reveals that both the decline in BE-passives and the increase in GET-passives spread from less-formal to more-formal genres. The most recent, speech-related Soap Opera corpus data are strikingly different from the other material as regards frequency of both kinds of passive, and gives some indication of how differently the constructions are probably being used in informal, spoken language compared to written language of any formality level.

Changes in the use of the passives are interpreted as possible manifestations of prescriptivism, colloquialization, and grammaticalization. The first two of these processes may play a role in the decreased use of the BE-passive, and the last two are likely playing some role in the increased use of the GET-passive. It was hypothesized that the double-proscription on passive voice and preposition stranding might have a visible effect on the diachronic use of the prepositional passive. However, in Article 4, the decreasing trajectory of the prepositional passive virtually mirrored that of the BE-passive, indicating that prepositional passives with BE may be responding more to the proscription on passive voice than on preposition stranding, which, along with the preposition-stranding findings in Article 1, offers further indication that some proscriptions are more salient and "successful" than others. A question I am interested in exploring here is what the findings in this thesis, taken together, suggest about the intertwined roles of prescriptivism, collo-

---

[44] The GET-passive frequencies from the Soap Opera Corpus are the more conservative extrapolated figures, excluding all instances of GET *married* (Schwarz 2015:163).

55

quialization, and grammaticalization in the use of the passive in American English.[45]

One very important concern in exploring the role of prescriptivism in language use is whether or not people are making deliberate choices in their use of language. That is, we can only claim that prescriptivism is behind the decline in BE-passives if writers are *intentionally and successfully* avoiding them. Language change can happen above the level of consciousness (that is, that people are aware of the meta-discourse surrounding the linguistic choice they are making) or below the level of consciousness (wherein people still select a linguistic form for a reason, but not consciously in response to linguistic meta-discourse) (see Curzan 2014: 58–60).[46]

The ability of prescriptivists to effect actual linguistic change is implied for the BE-passives in this thesis for two reasons: (1) the proscription on them is ubiquitous and vociferous and (2) the corpus language is written, edited, and published, which means it has been produced under circumstances very different from casual conversation and which imply a high level of meta-analysis. Interestingly, the decrease in BE-passives was also present in even the most recent, spoken-like material in the thesis (Article 1), which one would not necessarily expect. The decrease in passives in the Soap Opera Corpus both supports the idea that prescriptivist efforts have been "successful" in bringing about language change and casts doubt on this material's approximation of casual speech. However, it is difficult to the point of impossibility to find a way to empirically get at whether writers and editors in any of the corpora were aware of all of the passives they were using, especially in a diachronic study.

There are also a few arguments in favor of the "change from below the level of consciousness" hypothesis here. One very important one is the inability of the general public *and* the writers of style guides to even recognize a passive sentence (Pullum 2014: 73):

> The advice is often supplied by advice-givers who don't respect their own counsel – though they are unaware of that because they are commonly hopeless at distinguishing passives from actives. But the recipients of the advice can't identify passives either, so they are powerless to spot the blunders of their teachers.

If proscription is indeed having an effect on the use of BE-passives in written American English (as suggested by Leech et al 2009: 264), then writers *must* be at least marginally successful in recognizing it. A rather common view among linguists is that prescriptivist attitudes arise in response to language

---

[45] Again, it is recognized that claims about "American English" must be tentative due to the limited representativeness of the corpus material (section 3.1).

[46] These terms are not to be confused with other, sociolinguistic uses of "change from above" and "change from below," which refer to changes which spread from people with higher social status to people with lower social status ("from above") or vice versa ("from below").

change, but have limited (if any) effect on language use.[47] However, the use of certain written-language features does seem to be sensitive to widespread proscription, especially in the U.S., as in the *which*-hunt in restrictive relative clauses (Olofsson 2009; Leech et al. 2009: 264). Bohmann & Schultz (2011) and Hinrichs, Szmrecsanyi & Bohmann (2015) argue that the decline of *which* as a restrictive relativizer in American English must be attributed to this widespread prescriptive influence rather than to language-internal factors.

The prepositional passive with BE, as studied in Article 4, seems to constitute an example of a linguistic feature that is both sensitive *and* resistant to prescriptive influence. There are double proscriptions on the prepositional passive (i.e. *do not use passive voice* and *do not "strand" a preposition*); this means that we should expect to see a sharper decline in its frequency as compared to non-prepositional BE-passives if both proscriptions are effective. However, the frequency evidence in Article 4 suggests that prepositional BE-passives are declining at roughly the same rate as non-prepositional BE-passives. While this does not reveal anything about the "success" of prescriptivist efforts in reducing the number of BE-passives in written English, it does suggest that preposition stranding has been resistant to prescriptivist influence in the corpus material, as also found in the small-scale study of preposition stranding in Article 1. More research on preposition stranding in the corpora would obviously be needed in order to support this tentative conclusion.

Similarly difficult to prove, and less credible based on the data, is the idea that the proscription on GET has slowed the rise of the GET-passive (Anderwald 2016: 220, 245). For GET-passives, colloquialization and grammaticalization would seem to be driving a rise in frequency such that proscriptions, even if writers are willing and able to adhere to them, are having little effect.[48] Articles 2 and 3 in this thesis offer evidence, both in terms of increased frequency and along qualitative parameters, that the GET-passive is undergoing further grammaticalization. If so, then the signs of gradient change observed in the corpus data probably reflect earlier changes to the spoken language which are merely spreading through written genres, one formality level at a time. As GET-passives become more frequent and less marked in spoken language, the construction may be becoming less salient to writers and managing to "sneak" into the written language below the level of consciousness, especially in the more recent data.

In fact, as hard as it is to prove whether writers are deliberate in their use of passive voice, there is one rather good piece of evidence in the corpus data on the GET-passive that the spread of the construction is happening be-

---

[47] This view is problematized by Leech et al. (2009: 263) and at length by Curzan (2014).
[48] Also, it is impossible for us to know what the frequencies would have looked like had these features never been proscribed.

low the level of consciousness. Lieselotte Anderwald suggested that semi-passives should be taken into account when researching the development of the GET-passive (personal communication, August 23, 2016). Her suggestion led to the insight, in Article 3, section 3.1, that semipassives with GET are *not* actually increasing at anything like the rate that central GET-passives are. That is, it would seem to be *functional GET as a true passive auxiliary* that is spreading, and not the copula + adjective use from which it was derived (Fleisher 2006; see section 2.2.3). If it were the case that writers had been responding to proscriptions on GET, we might expect both constructions to be equally "slowed," as Anderwald suggests (2016: 220, 245).

We can reasonably assume that prescriptivism may affect language use if people are aware of their linguistic choices, and that grammaticalization is probably a kind of change that operates below the level of consciousness. Colloquialization, however, is a more difficult process to place along the deliberate/accidental usage divide. Colloquialization *only* applies to written language; this is language that is produced under very different circumstances from the "linguistic baseline" (Mair 2006: 183) of casual, face-to-face, spoken interaction.

In section 2.2.2, I explained how I consider democratization to be the force behind colloquialization of written genres. The question that remains to be explored regarding the BE-passive is whether democratization of discourse can be said to be behind the diachronic shifts in the use of passives in American English. Insofar as the exclusion of BE-passives and inclusion of GET-passives signal colloquialization, the answer must be "yes." However, prescriptive influence on the BE-passive also suggests a somewhat contradictory, anti-democratization mechanism at work.

For the sake of argument, let us say that the BE-passive is mostly declining due to colloquialization (Leech et al. 2009: 244). If so, then the decline must be due to democratization (the force behind colloquialization) and thus reflect increasing equality in society (optimistically) or the influence of market forces (rather more pessimistically: see section 2.2.2). If, however, as Leech et al. (2009: 264) *also* suggest, the BE-passive is declining in response to proscription, then democratization is likely playing a lesser role. While prescriptivism can reflect the influence of democratization in examples like official organizations adopting non-sexist language policies, the proscriptions on the BE-passive would *not* seem to reflect a move toward greater equality. Pullum (2010: 13) writes that style advice of this kind damages people's self-confidence and persuades them that their writing is bad; this is hardly synonymous with Fairclough's (1992: 201) "removal of inequalities and asymmetries in the discursive and linguistic rights, obligations and prestige of groups of people." Rather, the somewhat condescending advice of self-appointed linguistic gatekeepers cannot reasonably be viewed as a democratic effort to promote equality, but rather as a means of maintaining an

imbalanced power structure wherein what constitutes "good" English is the purview of a privileged few.

The increased use of the GET-passive is supposed to be due to both colloquialization and grammaticalization. As I have shown above, the corpus evidence in this study strongly supports grammaticalization as a factor in its increase. If the GET-passive is also increasing due to colloquialization, as Leech et al. (2009: 244) reasonably claim, then democratization is again necessarily the force behind the change. These are not conflicting claims. Grammaticalization is likely to start in spoken language and spread to written language in such a manner that its spread cannot be untangled from colloquialization. Therefore, increased use of this spoken-language feature in written genres, whose spread is *not* convincingly limited by prescriptivist efforts, is also taken to be an indicator of democratization, as its use helps make written language more accessible and informal.

In this section, I have reflected on the way theories of language change relate to the use of passive voice based on the evidence in my thesis. It is largely impossible to ever definitively determine which processes have had the greatest influence on the diachronic development of the use of BE-, GET- and prepositional passives. Another way of approaching research on linguistic change would of course be to start from the theoretical concept (prescriptivism, colloquialization, grammaticalization) and look for linguistic features that might signal that it is at play in language use. In such an endeavor, the interwoven nature of these processes should be taken into account, as the use of linguistic features may be influenced by *multiple* processes, as this thesis suggests for the use of passives.

7. Conclusion

This introductory survey has provided an overview not only of the separate articles, but of how the studies can be interpreted together to suggest new insights into the changing use of passive voice in American English. Again, findings cannot be generalized to an entire population of speakers; but corpus methodology does provide clues to the pathways of diachronic change and indicates how acceptable Americans have found the constructions to be over time.

In this survey, I discussed the background on the passives and the theories of language change that informed the topics of the thesis. I have also attempted to demonstrate why corpus investigation was deemed the best method for exploring changes in the passive in American English, and why the BYU corpora were selected as the primary sources of data, despite their limitations.

The classification systems that were developed for the studies were offered as a possible contribution to future work. These classification systems

were discussed and problematized, and refinements and revisions were suggested where appropriate.

Finally, the findings of the thesis overall were discussed. This entailed an overview of the diachronic development of the different constructions across corpora and genres, and of the differences between the kinds of passive in focus in the thesis. Furthermore, the greater question of *why* these changes have come about was considered at length in this survey. Considering the possible effect of prescriptivism, colloquialization, and grammaticalization on the passive constructions sheds light on how these three processes are intertwined.

While this thesis may offer some new insights into the development of three passive constructions in American English, there are certainly many more avenues to explore. Although corpora of spoken conversation tend to be quite small, it would be well worth tracing the use of the GET-passive through spoken language insofar as it is possible to do so with the digital resources currently available. It is likely that the GET-passive may show further signs of grammaticalization in spoken-language data.

More empirical work on the characteristics of prepositional passives is also called for. The frequency of [verb + preposition] can also be examined in the corpus overall, to give an idea of the role of overall frequency of use in encouraging passivization. Affectedness of the passive subject-referent can be further investigated by comparing results with thematic roles in active sentences and by considering the role of telicity in conferring affectedness on a subject-referent (Hopper & Thompson 1980; Croft 1994). Empirical use of the thematic role framework could probably be applied to other research questions as well, and it is likely that this framework can be refined further based on my exploratory attempt in Article 4.

The development of the prepositional passive along the parameters that have been studied here could also be traced back further in time in order to see whether it has always had the same semantic and discourse-structural preferences, or if these have developed over time. Cross-varietal, cross-genre studies of all of the constructions in the thesis can also be expected to be very fruitful, especially studies of emerging varieties of English across the world.

Finally, the intertwined roles of prescriptivism, colloquialization, and grammaticalization, which have begun to be the focus of a great deal of research, are well worth investigating further. Passive voice is only one of the linguistic features which are susceptible to the influence of multiple mechanisms, and there are many other ways of exploring language development along these three lines using a variety of linguistic features. Passive voice in American English will also continue to be an interesting feature to investigate, as its use in published writing offers revelations into attitudes towards language use and democratization in those language contexts: this must be looked into.

60

Corpora used in the thesis

Davies, Mark. 2004. *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

Davies, Mark. 2007. *TIME Magazine Corpus: 100 million words, 1920s–2000s*. Available online at http://corpus.byu.edu/TIME/.

Davies, Mark. 2008. *The Corpus of Contemporary American English: 450 million words, 1990–present*. Available online at http://corpus.byu.edu/COCA/.

Davies, Mark. 2010. *The Corpus of Historical American English: 400 million words, 1810–2009*. Available online at http://corpus.byu.edu/coha/.

Davies, Mark. 2012. *The Corpus of American Soap Operas: 100 million words, 2001–2012*. Available online at http://corpus.byu.edu/SOAP/.

Corpora referenced in the introductory survey

The Bank of English (BoE), sponsored by HarperCollins Publishers, Glasgow, and compiled by the COBUILD team at the University of Birmingham, UK.

The British National Corpus (BNC), BNC Consortium, Oxford University Press.

Brown Corpus, see *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown)*

BYU-BNC. See the above list of corpora used in the thesis.

The Corpus of American Soap Operas: 100 million words, 2001–2012. See the above list of corpora used in the thesis.

The Corpus of Contemporary American English: 450 million words, 1990–present. See the above list of corpora used in the thesis.

A Corpus of English Dialogues 1560–1760. 2006. Compiled under the supervision of Merja Kytö (Uppsala University) and Jonathan Culpeper (Lancaster University).

The Corpus of Historical American English: 400 million words, 1810–2009. See the above list of corpora used in the thesis.

The Freiburg-Brown Corpus ('Frown') (original version) Christian Mair (project leader), Albert-Ludwigs-Universität Freiburg.

The Freiburg-Brown Corpus ('Frown') (POS-tagged version) Christian Mair (project leader), Albert Ludwigs-Universität Freiburg, and Geoffrey Leech, Lancaster University.

The Freiburg-LOB Corpus ('F-LOB') (original version) Christian Mair (project leader), Albert-Ludwigs-Universität Freiburg.

The Freiburg-LOB Corpus ('F-LOB') (POS-tagged version) Christian Mair (project leader), Albert Ludwigs-Universität Freiburg, and Geoffrey Leech, University of Lancaster.

The International Corpus of English (ICE), University College London, Sidney Greenbaum (former project leader). See individual corpus manuals for compilers.

The Lancaster-Oslo/Bergen Corpus (LOB Corpus), original version (1970–1978), Geoffrey Leech, Lancaster University (project leader), Stig Johansson, University of Oslo (project leader), and Knut Hofland, University of Bergen (head of computing).

The Lancaster-Oslo/Bergen Corpus (LOB Corpus), POS-tagged version (1981–1986), Geoffrey Leech, Lancaster University (project leader), Stig Johansson,

University of Oslo (project leader), Roger Garside, Lancaster University, and Knut Hofland, University of Bergen (heads of computing).

LOB Corpus, see *The Lancaster-Oslo/Bergen Corpus (LOB Corpus)*.

London-Lund Corpus of Spoken English. Compiled by Jan Svartvik.

A Representative Corpus of Historical English Registers (ARCHER), version 3.2 (1990–2013). First constructed by Douglas Biber and Edward Finegan.

A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown). 1964, 1971, 1979. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island.

TIME Magazine Corpus: 100 million words, 1920s–2000s. See the above list of corpora used in the thesis.

Works cited in the introductory survey

Aarts, Bas. 2007. *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.

Abels, Klaus. 2003. Successive cyclicity, anti-locality and adposition stranding. Doctoral dissertation, University of Connecticut.

Anderwald, Lieselotte. 2016. *Language Between Description and Prescription: Verbs and Verb Categories in Nineteenth-Century Grammars of English*. Oxford: Oxford University Press.

Barber, E. J. W. 1975. Voice – beyond the passive. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* 1, 16–24.

Beavers, John. 2011. On affectedness. *Natural Language and Linguistic Theory* 29, 335–370.

Biber, Douglas. 1993a. Representativeness in corpus design. *Literary and Linguistic Computing* 8(4), 243–257.

Biber, Douglas. 1993b. Using register-diversified corpora for general language studies. *Computational Linguistics – Special Issue on Using Large Corpora II* (19)2, 219–241.

Biber, Douglas. 1995. *Dimensions of Register Variation*. Cambridge: Cambridge University Press.

Biber, Douglas & Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65(3), 487–517.

Biber, Douglas & Edward Finegan. 1992. The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New Methods and Interpretations in Historical Linguistics*, 688–704. Berlin & New York: Mouton de Gruyter.

Biber, Douglas & Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In Terttu Nevalainen & Leena Kahlas-Tarkka (eds.), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, 253–276. Helsinki: Société Néophilologique.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London & New York: Longman.

Bohmann, Axel & Patrick Schultz. 2011. Sacred that and wicked which: Prescriptivism and change in the use of English relativizers. *Proceedings of the Nineteenth Annual Symposium About Language and Society–Austin, Texas Linguistics Forum 54*, 88–101.

Bohner, Gerd. 2001. Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim. *British Journal of Social Psychology* 40, 515–529.

Bolinger, Dwight. 1975. On the passive in English. In Adam Makkai & Valerie Becker Makkai (eds.), *The First LACUS Forum 1974*, 57–77. Columbia, SC: Hornbeam Press.

Bolinger, Dwight. 1977. Transitivity and spatiality: The passive of prepositional verbs. In Adam Makkai, Valerie Becker Makkai & Luigi Heilmann (eds.), *Linguistics at the Crossroads*, 57–78. Padova, Italy: Liviana Editrice, and Lake Bluff, Illinois: Jupiter Press.

Bosse, Solveig, Benjamin Bruening & Masahiro Yamada. 2012. Affected experiencers. *Natural Language and Linguistic Theory* 30(4), 1185–1230.

Brinton, Laurel J. 1988. *The Development of English Aspectual Systems: Aspectualizers and Post-Verbal Particles*. Cambridge: Cambridge University Press.

Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Cameron, Deborah. 1995. *Verbal Hygiene*. London & New York: Routledge.

Carter, Ronald & Michael McCarthy. 1999. The English *get*-passive in spoken discourse: Description and implications for an interpersonal grammar. *English Language and Linguistics* 3(1): 41–58.

Chappell, Hillary. 1980. Is the *get*-passive adversative? *Papers in Linguistics: International Journal of Human Communication* 13(3), 411–452.

Collins, Peter C. 1996. *Get*-passives in English. *English World-Wide* 15(1), 43–56.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.

Coto Villalibre, Eduardo. 2015. Is the *get*-passive really that adversative? *Miscelánea: A Journal of English and American Studies* 51, 13–30.

Couper-Kuhlen, Elizabeth. 1979. *The Prepositional Passive in English: A Semantic-Syntactic Analysis, with a Lexicon of Prepositional Verbs* (Linguistische Arbeiten 81). Tübingen: Niemeyer.

Croft, William. 1994. Voice: Beyond control and affectedness. In Barbara A. Fox & Paul J. Hopper (eds.), *Voice: Form and Function*, 89–117. Amsterdam & Philadelphia: Benjamins.

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.

Curzan, Anne. 2014. *Fixing English: Prescriptivism and Language History*. Cambridge: Cambridge University Press.

Davies, Mark. 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25(4), 447–464.

Davies, Mark. 2016. Some methodological issues related to corpus-based investigations of recent syntactic changes in English. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 157–174. Oxford: Oxford University Press.

Davison, Alice. 1980. Peculiar passives. *Language* 56, 42–66.

Deignan, Alice & Liz Potter. 2004. A corpus study of metaphors and metonyms in English and Italian. *Journal of Pragmatics* 36, 1231–1252.

Denison, David. 1993. *English Historical Syntax: Verbal Constructions*. London & New York: Longman.

Denison, David. 1999. Gradience and linguistic change. In Laurel J. Brinton (ed.), *Historical Linguistics 1999: Selected Papers from the 14th International Con-*

ference on Historical Linguistics, Vancouver, 9–13 August 1999, 119–144. Amsterdam & Philadelphia: Benjamins.

Downing, Angela. 1996. The semantics of *get*-passives. In Ruqaiya Hasan, Carmel Cloran & David Butt (eds.), *Functional Descriptions: Theory in Practice*, 179–205. Amsterdam & Philadelphia: Benjamins.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3), 547–619.

Dreschler, Gea. 2015. Passives and the loss of verb second: A study of syntactic and information-structural factors. Doctoral dissertation, Radboud University, Nijmegen. Utrecht: LOT.

Engdahl, Elisabet & Anu Laanemets. 2015. Prepositional passives in Danish, Norwegian and Swedish: A corpus study. *Nordic Journal of Linguistics* 38(3), 285–337.

Fairclough, Norman. 1992. *Discourse and Social Change*. Cambridge & Malden: Polity.

Farrelly, Michael & Elena Seoane. 2012. Democratization. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 392–401. Oxford: Oxford University Press.

Findlay, Jamie Y. 2016. The prepositional passive in Lexical Functional Grammar. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-Driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*, 255–275. Stanford, CA: CSLI Publications.

Fleisher, Nicholas. 2006. The origin of passive *get*. *English Language and Linguistics* 10(2), 225–252.

Francis, W. Nelson. 1979. Problems of assembling and computerizing large corpora. In Henning Bergenholtz & Burkhard Schaeder (eds.), *Empirische Textwissenschaft: Ausbau und Auswertung von Text-Corpora*, 110–123. Königstein: Scriptor.

Garey, Howard B. 1957. Verbal aspect in French. *Language* 33(2), 91–110.

Givón, Talmy & Lynne Yang. 1994. The rise of the English *get*-passive. In Barbara Fox & P.J. Hopper (eds.), *Voice: Form and Function*, 119–149. Amsterdam & Philadelphia: Benjamins.

Goh, Gwang-Yoon. 2001. The advent of the prepositional passive: An innovation of Middle English? *English Studies* 82(3), 203–217.

Granath, Solveig. 2009. What's so good about a "good question"? A quantitative approach to 'modifier + question' in spoken American news discourse. In Rhonwen Bowen, Mats Mobärg & Sölve Ohlander (eds.), *Corpora and Discourse–and Stuff: Papers in Honour of Karin Aijmer*, 47–58. University of Gothenburg: Acta Universitatis Gothoburgensis.

Hinrichs, Lars, Benedikt Szmrecsanyi & Axel Bohmann. 2015. *Which*-hunting and the Standard English relative clause. *Language* 91(4), 806–836.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2), 251–299.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization* (7th edn.). Cambridge: Cambridge University Press.

Huddleston, Rodney. 1976. Some theoretical issues in the description of the English verb. *Lingua* 40(4), 331–383.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hundt, Marianne. 2001. What corpora can tell us about the grammaticalisation of voice in *get*-constructions. *Studies in Language* 25(1), 49–88.

Hundt, Marianne. 2007. *English Mediopassive Constructions: A Cognitive, Corpus-Based Study of their Origin, Spread, and Current Status*. Amsterdam & New York: Rodopi.

Hundt, Marianne & Geoffrey Leech. 2016. "Small is beautiful": On the value of standard reference corpora for observing recent grammatical change. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 175–188. Oxford: Oxford University Press.

Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *The International Journal of Corpus Linguistics* 4(2), 221–242.

Koopman, Hilda. 1984. *The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar*. Dordrecht, NL: Foris.

Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Leech, Geoffrey. 2004. Recent grammatical change in English: Data, description, theory. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in Corpus Linguistics*, 61–81. Amsterdam & New York: Rodopi.

Leech, Geoffrey. 2007. New resources, or just better old ones? The Holy Grail of representativeness. In Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.), *Corpus Linguistics and the Web*, 133–150. Amsterdam & New York: Rodopi.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

Lindquist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

Lippi-Green, Rosina. 1997. *English with an Accent: Language, Ideology, and Discrimination in the United States*. London & New York: Routledge.

Mair, Christian. 1997. Parallel corpora: A real-time approach to the study of language change in progress. In Magnus Ljung (ed.), *Corpus-based Studies in English*, 195–209. Amsterdam & New York: Rodopi.

Mair, Christian. 2004. Corpus linguistics and grammaticalisation theory: Statistics, frequencies, and beyond. In Hans Lindquist & Christian Mair (eds.), *Corpus Approaches to Grammaticalization in English*, 121–150. Amsterdam & Philadelphia: Benjamins.

Mair, Christian. 2006a. *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.

Mair, Christian. 2006b. Tracking ongoing grammatical change and recent diversification in Present-Day Standard English: The complementary role of small and large corpora. In Antoinette Renouf & Andrew Kehoe (eds.), *The Changing Face of Corpus Linguistics*, 355–376. Amsterdam & New York: Rodopi.

Mair, Christian. 2011. Grammaticalization and corpus linguistics. In Heiko Narrog & Bernd Heine (eds.), *The Oxford Handbook of Grammaticalization*, 239–250. Oxford: Oxford University Press.

Mair, Christian & Marianne Hundt. 1995. Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress. *Zeitschrift für Anglistik und Amerikanistik* 43(2), 111–122.

McEnery, Tony & Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Mitchell, Bruce & Fred C. Robinson. 2012. *A Guide to Old English* (8th edn.). Hoboken, NJ: Wiley-Blackwell.

Mukherjee, Joybrato. 2006. Corpus linguistics and English reference grammars. In Antoinette Renouf & Andrew Kehoe (eds.), *The Changing Face of Corpus Linguistics*, 337–354. Amsterdam & New York: Rodopi.

Nilsson [Schwarz], Sarah. 2012. Colloquialization in contemporary American English: The GET-passive and negative contractions in the TIME Magazine Corpus. Unpublished master's thesis.

Nilsson [Schwarz], Sarah. 2013. The representativeness problem. Unpublished term paper.

OED. 2011. *Oxford English Dictionary Online*. Oxford University Press. (www.oed.com)

Olofsson, Arne. 2009. Half a century of relative monitoring: On the 1959 which-hunt in Scientific American. *Moderna Språk* 102(2), 3–11.

Palmer, Frank Robert. 1965. *A Linguistic Study of the English Verb*. London: Longman.

Persson, Gunnar. 1990. *Meanings, Models and Metaphors: A Study in Lexical Semantics in English* (Umeå Studies in the Humanities 92). Stockholm: Almqvist and Wiksell.

Pullum, Geoffrey K. 2010. The land of the free and The Elements of Style. *English Today* 26(2), 34–44.

Pullum, Geoffrey K. 2014. Fear and loathing of the English passive. *Language and Communication* 37(7), 60–74.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Rodríguez-Puente, Paula. 2016. Tracking down phrasal verbs in the spoken language of the past: Late Modern English in focus. *English Language and Linguistics* 21(1), 69–97.

Saeed, John I. 2003. *Semantics* (2nd edn.). Oxford: Blackwell.

Schwarz, Sarah. 2015. Passive voice in American soap opera dialogue. *Studia Neophilologica* 87(2), 152–170.

Schwarz, Sarah. 2017. "Like getting nibbled to death by a duck": Grammaticalization of the GET-passive in the TIME Magazine Corpus. *English World-Wide* 38(3), 305–335.

Schwarz, Sarah. Under review a. Signs of grammaticalization: Tracking the GET-passive through COHA.

Schwarz, Sarah. Under review b. *This must be looked into*: A corpus study of the prepositional passive.

Seoane Posse, Elena. 1999. The consolidation of the indirect and prepositional passive in Early Modern English: Evidence from the Helsinki Corpus. *Estudios Ingleses de la Universidad Complutense* 7, 119–139.

Siemund, Rainer. 1995. 'For Who the Bell Tolls': Or why corpus linguistics should carry the bell in the study of language change in Present-day English. *Arbeiten aus Anglistik und Amerikanistik* 20, 351–376.

Smitterberg, Erik. 2005. *The Progressive in 19th-Century English: A Process of Integration*. Amsterdam & New York: Rodopi.

Stephens, Bret. 2017, August 25. Tips for aspiring op-ed writers. *The New York Times*. Retrieved from http://www.nytimes.com.

Strunk Jr., William & E. B. White. 2000. *The Elements of Style* (4th edn.). New York: Allyn and Bacon.

Svartvik, Jan. 1966. *On Voice in the English Verb*. The Hague & Paris: Mouton & Co.

Sussex, Roland. 1982. A note on the *get*-passive construction. *Australian Journal of Linguistics* 2, 83–92.

Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2), 143–160.

Wolf, Naomi. 2015, July 24. Young women, give up the vocal fry and reclaim your strong female voice. *The Guardian*. Retrieved from http://www.theguardian.com.

Yáñez-Bouza, Nuria. 2015. *Grammar, Rhetoric and Usage in English: Preposition Placement 1500–1900*. Cambridge: Cambridge University Press.